ARE TEACHER EXPERIENCE AND EDUCATION ASSOCIATED WITH EVALUATION RATINGS?

A Dissertation

Presented to

the School of Education

Drake University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

by Sherie K. Surbaugh

February 1993

ARE TEACHER EXPERIENCE AND EDUCATION

ASSOCIATED WITH EVALUATION RATINGS?

by Sherie K. Surbaugh

February 1993

Approved by Committee:

Dr. Hilda L. Williams, Chair

Dr. James L. Romig

Dr. Susan E. Wright

Dr. Richard L. Schwab
Dean of the School of Education

TABLE OF CONTENTS

Chapter		Page
1.	INTRODUCTION	1
	Statement of the Problem	5
	Significance of the Study	6
	Limitations	7
	Delimitations	8
	Assumptions	9
	Definition of Terms	9
	Summary	11
2.	REVIEW OF THE LITERATURE	13
	Introduction	13
	Review of the Research	13
	Summary	30
3.	DESIGN OF THE STUDY	33
	Setting and Population	33
	Instrumentation	34
	Data Collection Procedures	41
	Research Design	45
	Research Hypotheses	45
	Treatment of the Data	48
4.	ANALYSIS OF THE DATA	51
	Descriptive Data: Population	51

Page

Item Analysis 52

Descriptive Data: High-performance and Low-performance Groups 54

Hypotheses of the Study 56

Evaluation Ratings by Section of Evaluation Instrument 60

Section Ratings and Teacher Experience 61

Section Ratings and Formal Education 66

Descriptive Data: Grouping by Other Variables - Population 71

Descriptive Data: Grouping by Other Variables - High-performance and Low-performance Groups 76

Summary of Chapter 4 79

5. SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS 82

Summary of the Study 82

Conclusions 84

Recommendations 89

REFERENCES 91

APPENDICES

A. PREMISES FOR AN EVALUATION SYSTEM FOR TEACHERS 95

B. CRITERIA AND PHILOSOPHIC PREMISES 97

C. SUMMATIVE REPORT OF THE TEACHER EVALUATION INSTRUMENT 100

D. TEACHER EVALUATION INSTRUMENT WITH LEVELS OF PERFORMANCE 101

Page

E. CODES USED IN DATA COLLECTION 106

F. TOTAL EVALUATION RATINGS BY FREQUENCY AND PERCENTAGE: POPULATION 107

G. MEAN EVALUATION RATINGS BY YEAR EVALUATED: POPULATION 108

H. MEAN EVALUATION RATINGS, EXPERIENCE AND FORMAL EDUCATION: POPULATION 109

I. *P* VALUES AND DISCRIMINATION VALUES BY ITEM BASED ON "EXCEEDS" RATING 110

J. FREQUENCY OF EVALUATION RATINGS: HIGH-PERFORMANCE GROUP 111

K. FREQUENCY OF EVALUATION RATINGS: LOW-PERFORMANCE GROUP 112

L. MEAN EVALUATION RATINGS BY YEAR EVALUATED: HIGH- AND LOW-PERFORMANCE GROUPS 113

M. MEAN EVALUATION RATINGS, EXPERIENCE AND FORMAL EDUCATION: HIGH- AND LOW-PERFORMANCE GROUPS 114

N. EVALUATION SCORES BY SECTION: POPULATION ... 115

O. EVALUATION SCORES BY SECTION: HIGH- AND LOW-PERFORMANCE GROUPS 116

P. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY TEACHER GENDER 117

Q. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY SUBJECT TAUGHT 118

R. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY BUILDING 119

S. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY EVALUATOR 120

Page

T. RANKING OF EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY EVALUATOR 121

U. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY GENDER: HIGH- AND LOW-PERFORMANCE GROUPS 122

V. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY SUBJECT TAUGHT: HIGH- AND LOW-PERFORMANCE GROUPS 123

W. MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY BUILDING: HIGH- AND LOW-PERFORMANCE GROUPS 124

X. NUMBER OF CASES AND MEAN RATINGS BY EVALUATOR: HIGH- AND LOW-PERFORMANCE GROUPS 125

LIST OF TABLES

Table		Page
1.	Experience: High- and Low-performance Groups	59
2.	Formal Education: High-and Low-performance Groups	60
3.	Section 1 Evaluation Ratings by Experience	62
4.	Section 2 Evaluation Ratings by Experience	63
5.	Section 3 Evaluation Ratings by Experience	64
6.	Section 4 Evaluation Ratings by Experience	65
7.	Section 5 Evaluation Ratings by Experience	66
8.	Section 1 Evaluation Ratings by Formal Education	67
9.	Section 2 Evaluation Ratings by Formal Education	68
10.	Section 3 Evaluation Ratings by Formal Education	69
11.	Section 4 Evaluation Ratings by Formal Education	70
12.	Section 5 Evaluation Ratings by Formal Education	71
13.	Relationship between Section Evaluation Ratings and Experience	80
14.	Relationship between Section Evaluation Ratings and Formal Education	81
15.	Comparison of Mean Evaluation Ratings by Experience Group: Population	87

Page

16. Comparison of Total Evaluation Ratings by Formal Education: Population 88

LIST OF FIGURES

Figure — Page

1. Item analysis of evaluation instrument based on *Exceeds* ratings: High- and low-performance groups 54

Chapter 1

INTRODUCTION

Public education is under siege in America. Leaders in business, industry, government, and education are calling for wholesale changes in American schools. Al Shanker, President of the American Federation of Teachers, predicted that public education as we know it will not last more than 5 to 10 years if educators do not make fundamental changes (1988). President Bush held an Education Summit in 1989, only the second time a United States president had met with the nation's governors to discuss education issues (Lepley, 1989).

The current call for school reform began in 1983 with A Nation at Risk. At least 38 other reports were critical of public schools. A Nation at Risk warned that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (The National Commission on Excellence in Education, 1983, p. 3). Reports made over 250 recommendations about principals, teachers, curriculum, equity, school-business partnerships, parent involvement, funding, school organization, instruction, and support services (Northwest Regional Educational Laboratory, 1987).

More recent reports have continued to criticize public schools. U.S. Education Secretary Lauro Cavazos said he was "disappointed in the performance of education nationally" and called student performance in this country "stagnant" (Iowa Association of School Boards, 1989, p. 1). A report by the Center for Policy Research in Education contained "doubts" about the "rigor and challenge of new academic courses" after the era of education reform (cited in "Doubts Linger," 1989, p. 1).

Teacher performance and salary have been under scrutiny. The authors of A Nation at Risk (The National Commission on Excellence in Education, 1983) recommended that salary, promotion, tenure, and retention decisions "be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated" (p. 16). The authors also recommended that "career ladders distinguish among the beginning instructor, the experienced teacher, and the master teacher" and that contracts be extended to 11 months (The National Commission on Excellence in Education, 1983, p. 16). Other national reports recommended developing a national master teacher program, exploring approaches other than merit pay, rewarding teachers for outstanding performance, and offering

financial incentives for different levels of responsibility (Northwest Regional Educational Laboratory, 1987).

The quality of the teaching profession has been questioned. As early as 1983 Ernest Boyer warned that the teaching profession was weakened by a leveling off of salaries and a lack of a career ladder. He observed that to "get ahead" in teaching, a person had to leave the classroom (p. 7). The National Governors Association expressed concern about whether the teaching profession is able "to attract and keep able teachers." A report by The National Governors Association, Time for Results, contained 11 recommendations, including using career ladders, as a way to "redesign the structure of the teaching career by increasing levels of responsibility and compensation" (cited in Bobbitt, 1989, p. 10). A number of states established a Master's degree requirement in an attempt to improve teaching (Knapp, McNergney, Herbert, & York, 1990).

In 1984 an in-depth study of Iowa's education system was conducted by the Excellence in Education Task Force at the request of the Legislative Council. The Teaching Quality Subcommittee recommended that the Department of Public Instruction develop models for a state career ladder system in which salary would be "commensurate with the training, skills, and responsibilities required in the profession" (p. 36). Noting "negative experience and lack

of research on merit pay plans," the Subcommittee recommended "further study, research and experimentation into merit pay and other alternative methods of compensating educators" (p. 36).

In 1987 a task force assembled by The Iowa Association of School Boards recommended that school districts provide additional salary to teachers based on superior work or additional responsibility (Iowa Association of School Boards, 1987). The authors also recommended that assessment of teacher performance should be a continuous process and part of a comprehensive plan for professional growth.

The Iowa legislature enacted the Educational Excellence Program in 1987 to improve both the performance and salaries of Iowa's teachers. The goal of Phase III, funded at $41 million, was to enhance the quality, effectiveness, and performance of Iowa's teachers through performance-based and supplemental pay plans (State of Iowa, Chapter 294A, 1987).

Many Iowa school districts developed performance-based Phase III plans in which teachers could earn additional salary through meeting or exceeding district standards for performance evaluation. However, many plans were structured so that teachers with more experience and more formal education could earn the maximum amount while teachers with the same evaluation ratings but less experience and formal education could earn a lesser amount (Eckles, personal

correspondence, January, 1990). This practice continued the prevalent reward structure in which teachers with more experience and formal education received higher salaries than teachers with less experience and less formal education. Little or no data were available supporting or refuting the value placed on experience and formal education as criteria in Phase III performance-based pay plans (Eckles, personal correspondence, January, 1990).

Calls for change in public education and the teaching profession are widespread and persistent. Ways to improve teaching performance and the teaching profession are not readily apparent, and reports have advocated numerous approaches. This study is one response to the call to improve teaching performance and the teaching profession.

Statement of the Problem

In an effort to improve teaching performance and the teaching profession, school reform advocates have called for changes in the methods used to determine teacher salaries. Teacher salaries in most districts are determined by teacher experience and formal education. This practice has been widespread since the 1950s (Needham, 1982). Knienapfel (1984) reported that approximately 96% of school districts used teaching experience and formal education as the criteria for teacher salaries in the early 1980s. Teachers with the fewest years of experience and a Bachelor's degree

earn the least salary. Teachers with the most years of teaching experience and advanced degrees earn the highest salaries.

Teacher performance is typically evaluated by a building administrator on a regular schedule. As long as teacher performance generally meets the district's standards, teacher salary is determined by experience and formal education.

The purpose of this study was to determine the relationships between teaching performance and experience and formal education. Teaching performance was determined by evaluation ratings by building administrators. High-performance teachers were compared to low-performance teachers in experience and formal education to determine if evaluation ratings are positively associated with experience and formal education.

Significance of the Study

If teacher evaluation ratings are positively associated with teacher experience and formal education, then the present reward system would tend to be supported. If the high-performance teachers are those with more experience and formal education, then using experience and formal education as salary determination criteria would tend to be supported. If, however, low-performance teachers have as much experience and formal education as high-performance

teachers, then the present reward system would appear to be insupportable.

If teacher evaluation ratings are not associated with experience and formal education, then a continuation of the present reward system would be more difficult to justify. Results from this study may be useful in an assessment of the strengths and weaknesses of the present teacher reward system.

Limitations

First, the generalizability of the results to other districts may be limited due to the unique characteristics of the district under study. The district is described in Chapter 3.

A second limitation of the study is the definition of experience. District policy sets a cap on the maximum number of years of experience the district recognizes when a teacher is hired. For example, if a teacher were hired with 10 years of experience, the teacher would be credited with only 8 years of experience, the maximum. Experience in a setting outside a recognized public or private school would not be reflected in number of years of experience. Experience other than that recognized by the district may be influencing evaluation ratings more than is apparent in this study.

A third limitation of the study is the definition of formal education. Formal education is defined as formal education recognized by the school district. A teacher may participate in staff development training without receiving college credit, may have taken college credit courses without having received prior approval from the school district, may have earned a degree without approval or prior approval from the school district, or may have had other experiences which are not formally recognized as education by the school district. Education other than that recognized by the district may be influencing evaluation ratings more than is apparent in this study.

A fourth limitation of the study is an absence of current data about the validity of the ratings obtained from the evaluation instrument.

A fifth limitation of the study is that reliability data for the evaluation ratings used in this study are not available.

Delimitations

Due to the unique nature of the evaluation instrument, this study is limited to secondary teachers in selected subject areas in one school district.

Assumptions

The following assumptions are commonly made by school districts using instruments to evaluate teacher performance:

1. The evaluation instrument can be used to discriminate among degrees of effective teaching behaviors.
2. There is consistency among evaluators.
3. There is consistency within evaluators, i.e., the same standard applies to each teacher.
4. Each criterion in the evaluation instrument is equal in importance to the others.
5. Criteria comprising the evaluation instrument represent performance standards of the district.
6. Criteria used for the evaluation instrument are related to effective instruction.
7. The evaluator's ratings represent a valid measure of a teacher's performance.
8. Each appraiser followed all evaluation procedures, including conducting the pre-conference and the observation, completing the evaluation form, and meeting with the teacher in the post-conference.

Definition of Terms

Criteria are the characteristics, behaviors, or outcomes upon which judgments are made about a teacher's performance.

Education is the formal education of the teacher recognized by the district. Qualifying college courses, workshops, and degrees are determined by the school district. Prior approval must be received. This study grouped teachers into two categories by education: those who had not earned a recognized Master's degree and those who had.

Evaluation instrument is composed of the criteria upon which teachers were evaluated during the period of the study.

Evaluation system is the set of procedures used to evaluate teacher performance. The procedures include a pre-conference between the administrator and teacher, the classroom observation by the administrator, and the post-conference between the administrator and the teacher.

Evaluators are building-level administrators responsible for making periodic judgments about the performance of teachers.

Experience is the number of years of teaching experience the school district recognizes. School district policy provides guidelines for recognizing partial and part-time experience.

Formative evaluations are the observations and data gathered to promote improvement rather than to make personnel decisions. (See summative evaluations.)

Merit pay is a method of determining teacher salary based on periodic assessments of the teacher's performance.

Reliability is the extent to which teacher evaluation ratings are consistent across time.

Summative evaluations are the performance ratings upon which personnel decisions such as compensation, tenure, and re-employment are made.

Validity is the degree to which an evaluation instrument measures what it purports to measure.

Summary

The demands for school reform are strong and pervasive. A frequent call is to link teacher performance to salary. Presently, the years of teaching experience and amount of formal education determine salary. Although teachers are evaluated regularly, salary is not dependent upon evaluation ratings. As long as teachers meet the school district's minimum expectations on the evaluation instrument, salary increments are determined by years of experience and amount of formal education recognized by the district.

The purpose of this study is to determine if teacher experience and formal education are associated with evaluation ratings. The results of the study can be used in making policy decisions about restructuring the teaching profession. School reform efforts in Iowa, as well as in

other states, can be strengthened by examining present practice and basing reform efforts on a sound rationale.

Chapter 2

REVIEW OF THE LITERATURE

Introduction

Reports such as A Nation at Risk (The National Commission on Excellence in Education, 1983) and Time for Results (cited in Bobbitt, 1989) recommended revising teacher compensation systems to recognize performance. Most school districts use teaching experience and formal education as the criteria for salary determination rather than using measures of performance.

This review of literature begins with an historical description of salary determination methods prior to today's use of experience and formal education. The advantages and disadvantages of using experience and formal education are described. Although merit-pay plans have been supported in the past, controversy over teacher evaluation has prevented their continued success. Finally, research studies which attempted to determine the relationships between evaluation ratings and experience and formal education are discussed.

Review of the Research

Early Methods to Reward Teachers

Prior to the Civil War, teacher salaries were negotiated individually. Teacher salary was dependent upon

school board preferences and the teacher's negotiation skills rather than on teacher characteristics, teaching performance, or student learning (Lortie, 1975; Needham, 1982).

In the early twentieth century, teacher salary was determined by the grade-level taught, gender, role, and experience of teachers. Using Newark, New Jersey's practice in 1917 as an example, Needham (1982) reported that high school teachers earned twice that of elementary teachers; salaries for men were 20% to 25% higher than women in the same positions; some positions were open only to men; and a teacher was not entitled to additional salary after six years of experience. The teaching profession offered little incentive for self-improvement due to the combination of low salary and little expectation of anything better (Needham, 1982).

From 1900 to 1920, the number of male teachers declined. World War I offered teachers, especially women, the chance to obtain government jobs at up to twice the pay they had been receiving. To alleviate the resulting shortage in teacher supply, the National Education Association (NEA) Committee on the Emergency in Education recommended in 1920 that districts pay teachers "on the basis of education, professional training, and successful experience" (Needham, 1982).

Use of Experience and Education to Determine Salaries

The earliest report of experience and education being used to determine teacher salaries was in 1917 (Needham, 1982). Between 1920 and 1950, districts increasingly used experience and formal education as the factors to determine salaries (Tecker, 1985). According to a 1924 NEA survey, 20% of large school districts were using experience and education as the factors to determine salaries. By 1948, 95% of large school districts were using experience and education rather than level taught, responsibility, or gender (Needham, 1982). By far the most prevalent factors for teacher compensation in recent years have been a teacher's experience and education (Castetter, 1981; Center for Public Sector Labor Relations, 1985).

Although using experience and education to determine teacher salaries has been widespread, writers have pointed out its advantages and disadvantages.

Advantages in Using Experience and Education to Determine Salary

Using experience and education as criteria to determine teacher salaries offered these advantages:

1. Women can earn salaries equal to men (Needham, 1982).
2. Elementary teachers can earn salaries equal to secondary teachers (Needham, 1982).

3. Salary determination is easy to understand, administer, and utilize in budget preparation (Castetter, 1981).
4. By rewarding experience, the teaching force is stabilized (Needham, 1982).
5. By rewarding formal education, the teaching force is professionalized (Needham, 1982).
6. Teachers are paid according to objective standards, thereby reducing conflict and avoiding embarrassment (Lortie, 1975).

Although using experience and education as the criteria to determine teacher salaries reduced inequities, stabilized the teaching force, and was easy to administer, the practice has also been subject to criticism.

Disadvantages in Using Experience and Education to Determine Salaries

The use of experience and education as criteria for salary determination has been criticized on the following grounds:

1. The assumption that teacher effectiveness increases with experience and education may simply not be true (Castetter, 1981).
2. All teaching positions may not be equal in importance and responsibility (Castetter, 1981).

3. The practice fails to motivate teachers to achieve high levels of performance (Lortie, 1975).
4. Salary increases are virtually automatic and do not primarily depend on performance appraisal (Castetter, 1981).
5. The system encourages teachers to prepare for administrative positions (Lortie, 1975).
6. The power of the school board in determining salary is limited (Lortie, 1975).
7. Recruitment rather than retention has been favored (Lortie, 1975).
8. Administrative influence is limited since the teacher's salary is not dependent upon evaluation by the administrator (Lortie, 1975).
9. Since administrative influence is limited, teacher individualism has been encouraged (Lortie, 1975).
10. Using such rigid criteria as experience and education encourages teachers to leave education for fields with more attractive reward structures (Castetter, 1981).
11. Compensation is a function not of expertise and level of responsibility but of time on the job (Tecker, 1985).

12. Paying teachers according to their experience and education is likely to continue the preponderance of women in classroom teaching (Lortie, 1975).

Attempts to Pay Teachers According to their Performance

The call to revise teacher compensation systems to recognize performance is not a new idea. Travers (1981) reported a payment-by-results system begun in England in the late Victorian period. The system was based on the assumption that teachers, rather than students, were responsible for student achievement. Teacher salary was based on the results of annual student achievement tests administered by school inspectors. Travers reported that "the system corrupted the entire educational program" and these schools "came to represent the worst educational program of any civilized country" (p. 17).

In the United States, an early effort to link teacher performance to salary was begun in Kansas City, Missouri, in 1904. Teachers were required to pass yearly examinations in order to earn an annual salary increase. The testing program required that teachers increase their knowledge and skill in the theory and practice of education, as well as in history and philosophy (Guernsey, 1986).

Guernsey (1986) reported one school district in 1909 used teaching experience and instructional improvement to determine teacher salaries. In this merit pay plan,

teachers were able to pass through four stages which were very similar to career ladders. In moving through the stages, the teacher received substantial supervision, prepared and defended an extensive essay to the school board, completed a written examination, and completed college courses (Guernsey, 1986).

Castetter (1981) and Robinson (1983) reported on attempts in the United States to link teacher performance and salary, typically termed merit pay. After reaching its peak in the 1920s, merit pay decreased in the 1930s and 1940s as the use of experience and education increased. Knienapfel (1984) reported that interest in merit pay increased in the 1960s, when approximately 10% of school districts used a merit pay plan. By the late 1970s, however, use of merit pay plans again declined, when only 1% of districts used a merit or incentive pay plan (Porwoll, cited in Educational Research Service [ERS], 1988). The 1988 ERS Survey found that interest has increased again in the 1980s. By 1988 over 9% of responding school districts used their teacher evaluation results for incentive pay increments and another 20% used evaluation results for promotion in career ladder programs (Educational Research Service, 1988).

Merit pay plans typically based teacher salary on teaching experience, education, additional duties, and

performance evaluation (Tecker, 1985). Although using teaching experience and education as factors continued past practice, using extra work, harder work, or more important work added new factors to teacher salaries. The most controversial aspect of merit-pay plans, however, was using judgments about teaching performance as a factor in salary increases (Tecker, 1985).

Castetter (1981) and Robinson (1983) identified teacher evaluation as the primary cause of failure of merit pay plans. Researchers identified these specific reasons for the failure of merit pay plans:

1. Lack of knowledge of what constitutes quality instruction (Lortie, 1975)
2. Lack of ability to measure teaching performance, specifically the fairness, validity, and reliability of evaluation procedures (Lortie, 1975; Robinson, 1983)
3. Damaged teacher morale (Lortie, 1975; Robinson, 1983)
4. Administrative problems (Robinson, 1983)
5. Arbitrary distinctions (Robinson, 1983)
6. Inadequate financial incentives (Robinson, 1983)
7. Lack of definition of superior results (Robinson, 1983)
8. Inability to measure results (Robinson, 1983)

Tecker (1985) reported that the two teacher unions, the National Education Association (NEA) and the American Federation of Teachers, have opposed using factors other than experience and education to determine teacher salaries, arguing that excellence in teaching cannot be defined and rewarded (Tecker, 1985). The NEA has supported using experience and education arguing that they provide foreseeable increments and use the "two most objective factors related to teaching abilities" (NEA, 1985, p. 13).

The NEA identified these problems when evaluation of performance affects salary: (a) difficulty in determining who receives merit pay; (b) insufficient data to support evaluation; (c) subjective evaluation; and (d) inconsistency among evaluators (NEA, 1985).

In contrast to teachers' unions, teachers and the general public have expressed support for new methods of determining teacher salaries (Cornett, 1989; Tecker, 1985). In a national poll of teachers conducted by *The American School Board Journal*, over 60% of the statistically representative sample believed that effectiveness should be considered along with experience and education in determining their salary (Tecker, 1985). According to the Gallup Poll, more than 8 of 10 Americans favor salary increases for teachers who prove themselves particularly

capable. This is a higher percentage than the Poll showed five years ago (Cornett, 1989).

Castetter (1981) stated that much remains to be done to improve methods for appraising the performance of personnel, especially teaching effectiveness. He cautioned, however, it would be a mistake to assume that obstacles to rewarding performance are so great that they can never be overcome.

Teacher Evaluation

Nearly all educators agree that teacher evaluation is desirable (McGreal, 1983). According to McGreal, difficulties arise with evaluation not because of its general purposes, but from the way evaluation is carried out. According to the Educational Research Service (1988), a comprehensive teacher evaluation system should serve three major purposes:

1. To ensure that all teachers are at least minimally competent
2. To improve the performance of competent teachers
3. To identify and recognize the performance of outstanding teachers

Robinson (1983) and Castetter (1981) identified these characteristics as required in an effective evaluation system:

1. The criteria are significant and directly related to the organization's goals.

2. The same results are shown in different instances, e.g., the evaluation results are reliable.
3. The instrument measures what it purports to measure, e.g., the instrument is valid.
4. The evaluation is unbiased.
5. The evaluation procedures are practical.
6. The procedures are implemented by trained management.

Darling-Hammond, Wise, and Pease (1983) described seven methods to assess teacher competence, performance, or effectiveness: teacher interviews, competency tests, indirect measures, classroom observation, student ratings, peer review, student achievement, and faculty self-evaluation. Darling-Hammond et al. described teacher characteristics such as training and experience as examples of indirect measures which are "linked to teacher salary and promotion opportunities and, implicitly, to teacher evaluation" (p. 306).

Darling-Hammond et al. (1983) described teacher evaluation models which differed in their philosophical base, goals, and procedures. Since evaluation is value-laden and must be suitable for the situation, Darling-Hammond et al. emphasized that "any kind of evaluation activity involves value choices--and conflicts--at all levels of the operating system" (p. 312).

Travers (1981) cautioned that "there is no single simple method of evaluating teaching effectiveness, because there is no single concept of what the teacher should be undertaking in the classroom" (Travers, p. 14).

Finally, Darling-Hammond et al. (1983) described a defensible evaluation process as one that allows schools to balance these four goals: sort teachers, maintain staff morale, maintain organizational integrity, and promote incremental change. Epstein (1985) pointed out the need for fair and comprehensive evaluation standards regardless of whether the results are used for merit pay or for professional development.

Because of the difficulties experienced in implementing a merit pay plan, most school districts have continued with or returned to using experience and education as the criteria for salary determination. Few studies have been conducted, however, to determine the relationships between evaluation ratings and experience and formal education (Hoogeveen & Gutkin, 1986).

Relationship between Evaluation Ratings and Teacher Experience

Researchers have attempted to determine if a relationship exists between evaluation ratings and years of teaching experience. As the basis for evaluation ratings, researchers have used principals' ratings (Epstein, 1985;

Hoogeveen & Gutkin, 1986); teacher self ratings (Hoogeveen & Gutkin, 1986); principal ratings on a new statewide performance evaluation instrument (Micceri, 1984); and parent ratings (Epstein, 1985).

Using a sample of 44 teachers in three public midwestern elementary schools, Hoogeveen and Gutkin (1986) found no significant relationship at the .05 level between years of teaching experience and ratings by either principals ($r = .07$) or peers ($r = -.13$). Hoogeveen and Gutkin suggested that the results have implications for traditional methods of determining teacher salary since most school districts use teaching experience as a major factor. Hoogeveen and Gutkin concluded, "Taken together these data indicate that despite potential shortcomings, merit pay systems may still be more equitable than current practice" (p. 380).

Micceri (1984) found that years of teaching experience showed no relationship to effectiveness scores when the Florida Performance Measurement System was normed. Micceri noted a trend, "although statistically non-significant and of small magnitude," in a steady drop in mean evaluation scores as experience increased (p. 24). Micceri used five categories for years of experience and two categories for level taught.

Epstein (1985) questioned whether measures of teacher background, including experience and formal education, correlated with either principal or parent ratings. Epstein's study included 77 first-, third-, and fifth-grade teachers. A six-point scale from poor to outstanding was used to rate the teachers in classroom lessons, knowledge of subjects, discipline, and creativity. Epstein found that teacher experience was not a significant correlate of either principal or parent ratings.

Borich (1977) reported that experience variables have been "almost worthless in predicting teacher performance" but questioned "if the variables have been measured too grossly to yield significant findings" (p. 14).

In a study to determine the relationship between teacher experience and teacher effectiveness among secondary English teachers, Bastress (1980) used both total years of teaching experience and years of experience in the present school. Teacher effectiveness was measured by student responses on the Purdue Teacher Evaluation Scale. Bastress reported that students do not perceive more experienced secondary English teachers as being better able to structure the task or as having more position power.

In summary, researchers (Bastress, 1980; Borich, 1977; Epstein, 1985; Hoogeveen & Gutkin, 1986; Micceri, 1984) using a variety of teacher performance measures found no

relationship between evaluation ratings and years of teaching experience.

Relationship between Evaluation Ratings and Formal Education

Knapp, McNergney, Herbert, and York (1990) reported that research on the relationship between measures of teaching success and a teacher's formal education is rather limited. In addition, they reported that research is hampered by the lack of consensus about the meaning of the term "Master's degree" and the absence of a set of critical attributes to define the Master's degree for teachers (Knapp et al., 1990).

Micceri (1984) reported no significant relationship between scores on the Florida Performance Measurement System summative instrument and the highest educational degree attained by teachers. Trained observers rated over 1,200 elementary, middle school, and high school teachers in 41 schools in Florida. Micceri used four highest-degree-attained categories and two level-taught categories.

In a 1985 study of 77 first-, third-, and fifth-grade teachers, Epstein found that teachers with more credits beyond the Bachelor's degree were not necessarily viewed by principals or parents as better teachers than teachers with fewer accumulated credit hours. Teacher performance was measured using an instrument "similar to those used in several merit/master teacher plans" (p. 4). Epstein

suggested that neither the public nor the teaching profession will continue to support the lock-step salary schedules that pay the worst teachers as much as the best. Epstein warned that competent individuals will not enter a profession that continues to limit challenge, compensation, and advancement. In addition, Epstein suggested that the fairest evaluations use multiple judges, including parents, principals, and teacher-peers.

Bastress (1980) studied the relationship between teaching effectiveness and the formal education of secondary English teachers. Bastress used eight formal education categories from "Less than Bachelor's Degree" to "Doctoral Degree." Teacher effectiveness was measured by student responses on the Purdue Teacher Evaluation Scale. Bastress reported that students do not perceive secondary English teachers with more formal education as being able to structure the task better or as having more position power.

Knapp et al. (1990) summarized findings of Ashton, Crocker, and Olejnik (1986). One goal of the Ashton et al. study was to determine whether teachers with Master's degrees were more effective than those with Bachelor's degrees. Ashton et al. reviewed 15 studies, some of which used multiple samples and outcome measures. The studies used student achievement, teacher knowledge, or teacher performance to measure teacher effectiveness. Ashton et al.

reported conflicting results (Knapp et al., 1990). In seven studies no significant relationship between teachers' classroom performance or students' achievement and level of degree was found; however, in eight studies there was a statistically significant positive relationship (Ashton et al., 1986, as quoted in Knapp et al., 1990).

In summary, evidence of a relationship between teacher performance and formal education is mixed. Three researchers (Bastress, 1980; Epstein, 1985; Micceri, 1984) found no relationship between evaluation ratings and amount of formal education while Ashton et al. (1986) found a statistically significant relationship in 8 of 15 studies reviewed (Knapp et al., 1990).

International Comparisons

Barro and Suter (1988) studied the degree to which experience and education were the basis for determining salaries in eleven industrialized countries. They found that Japan, at one extreme, paid experienced teachers three times as much as beginning teachers. At the other extreme, Sweden and Denmark paid experienced teachers only 30% more than beginning teachers. The United States fell in the mid-range among the countries in rewarding teachers for increased teaching experience.

Barro and Suter (1988) also found that only France rewarded teachers on criteria other than experience and

education. France used an examination as the determiner for merit. Barro and Suter offered a number of questions for further study, including whether teaching performance is improved in countries which pay additional salary for education.

Summary

Using experience and education as the criteria to determine teacher salaries began in the early twentieth century to eliminate inequities, maintain a stable teaching profession, and encourage professional development (Needham, 1982). By the 1950s, most school districts used teacher experience and education to determine salaries (Needham, 1982).

Numerous attempts to link teacher performance to salary have been made (Castetter, 1981; Robinson, 1983). Merit-pay plans based on student performance proved damaging to the educational system in Victorian England (Travers, 1981). In the United States, merit pay plans gained popularity in the 1920s and the 1960s (Castetter, 1981; Knienapfel, 1984; Robinson, 1983). The most frequently mentioned reason for failure of merit-pay plans was conflict over teacher evaluation. During the 1980s, the school reform movement and public opinion favored performance-based salaries rather than salaries based on experience and formal education (Cornett, 1989; Tecker, 1985).

Nearly all educators agree that evaluation is desirable (McGreal, 1983). The difficulty is in relating evaluation to salary. Darling-Hammond et al. (1983) acknowledge that teacher evaluation is value-laden but suggest that a defensible evaluation process can allow schools to balance the goals of sorting teachers, maintaining staff morale, maintaining organizational integrity, and promoting incremental change.

Although Hoogeveen and Gutkin (1986) reported that relatively little research has been done to determine the relationship between evaluation ratings and experience, studies by Hoogeveen and Gutkin (1986), Borich (1977), Micceri (1984), and Epstein (1985) showed a lack of relationship between evaluation ratings and teaching experience.

Researchers have found no relationship between evaluation ratings and years of teaching experience (Bastress, 1980; Borich, 1977; Epstein, 1985; Hoogeveen & Gutkin, 1986; Micceri, 1984). No clear-cut evidence of a relationship between teacher performance and formal education is apparent (Ashton et al., 1986, as quoted in Knapp et al.). Barro and Suter (1988) recommended further study to determine if teaching performance is improved in countries which pay additional salary for formal education.

A review of the literature offers few clear answers. Using experience and education as criteria to determine salary has both advantages and disadvantages. Merit-pay plans have generally been unsuccessful. A small number of research studies report no relationship between teacher performance and experience. The findings of the research linking teacher performance and formal education are mixed.

Castetter (1981) pointed out that the assumption that teacher effectiveness increases with experience and education may simply not be true. This study is intended to provide data about the relationship among teacher effectiveness, experience, and formal education. To narrow the focus, this study included only teachers who had similar teaching assignments at the secondary level. This study can contribute to the research base needed to help make policy decisions to improve teaching and the teaching profession. A discussion of the methodology follows in Chapter 3.

Chapter 3

DESIGN OF THE STUDY

The purpose of this study was to determine the relationship of teacher experience and formal education to teacher performance as perceived by building administrators. This study used categorical data to determine if evaluation ratings were associated with experience or with formal education. In addition, evaluation ratings were grouped by the gender of the teacher, the subject taught, the building assignment, and the evaluator to investigate alternative explanations for the variability in evaluation ratings.

Setting and Population

This study was conducted in a suburban school district in the midwest. The community grew by over 20% from 1980 to 1988 (Iowa Department of Economic Development, 1991). During the period of this study, the school district consisted of one high school, two junior highs, and nine elementary buildings with an approximate student enrollment of 7,000. The district was among the largest 20 school districts in the state (Eckles, personal correspondence, July, 1992).

The initial population of this study was 86 full-time secondary school teachers. Subjects were limited to secondary language arts, math, social studies, and science

teachers to minimize differences in teaching roles and levels. Elementary teachers were not included since their teaching settings were different; counselors and nurses were not included since their evaluation criteria were different from classroom teachers.

The school district used a three-year evaluation cycle. Each teacher included in the study had been evaluated at least once during the 1988-89, 1989-90, or 1990-91 school year. If a teacher had been evaluated more than once, the most recent evaluation was used. Due to seven instances of a missing rating, an unclear rating, or a rating of "Not Observed," the population studied was reduced from 86 to 79 secondary teachers.

Instrumentation

Instrumentation consisted of the Teacher Performance Evaluation Instrument used by the district. The instrument was developed as part of a two-year project in 1973 to improve the performance evaluation system. The school district convened a Teacher Evaluation Commission comprised of teachers, administrators, and board members which later worked with consultants from Iowa State University to develop an instrument in which teachers' performance could be evaluated (Report on the Development of a Performance Evaluation System [RDPES], 1974). The Commission developed

eight premises for the evaluation system, three of which are quoted below:

1. A school community has the right to expect that the school district will conduct "a valid and continuous evaluation of the service of all district employees" (RDPES, 1974, p. 12).
2. "The essential purpose of evaluation is to improve performance by identifying an educator's strengths and weaknesses and to provide guides to develop a course of action to improve upon deficiencies" (RDPES, 1974, p. 13).
3. "It is possible to differentiate levels of performance" (RDPES, 1974, p. 13).

See Appendix A for the district's eight Premises for an Evaluation System for Teachers.

In selecting items for the evaluation instrument, the Teacher Evaluation Commission used "extensive statistical treatment, considerable bibliographic research, [and] massive sampling for the discrimination and validity testing" (RDPES, 1974, p. 2). Approximately 400 items were written by the Evaluation Commission and research team (RDPES, 1974, p. 19). These items were then linked to the 17 teacher roles identified by the Commission and research team. (See Appendix B for the Criteria and Philosophic Premises for the 17 teacher roles.) The remaining "138

items for teacher performance evaluation were then randomly placed in what were called 'Appropriateness and Observability' Survey Instruments" (RDPES, 1974, p. 20). These 138 items were then "submitted to a judgment panel of students, administrators, teachers, parents and board members from the school district to provide 'social validity' for the items" (RDPES, 1974, p. 20). Subsequently, the number of items was reduced to 82 (RDPES, 1974, p. 23). Each of the 82 items was then analyzed for discriminatory power (RDPES, 1974, p. 23). Seventy-six items met the discrimination index specifications (RDPES, 1974, p. 24). "The final step in the item selection process involved consideration of each item's appropriateness and discriminatory power" (RDPES, 1974, p. 24). Only items which had been rated "highly important" on a five-point scale on appropriateness or were basic premise descriptors, and had a discrimination index above 23% were returned (RDPES, 1974). "Using this approach, 76 items were chosen for the teacher performance evaluation pool" (RDPES, 1974, p. 25).

The following limitations were noted in the Report on the Development of the Performance Evaluation System:

1. "The process for determination of appropriateness and discrimination power was specifically tailored"

to the district and "its goals and values relative to desirable teacher performance" (p. 25).

2. "Students used in both the judgment panel for appropriateness and observability and in the evaluation of selected teachers" for item discrimination analysis "were predominately [sic] from the secondary grades" (p. 25).
3. The development of the instrument stopped "short of empirically testing the total system." "This step is recommended for the 1974-75 school term. Consequently, there has been no 'norming' experience with the total scale of the instrument" (p. 25).

The Executive Director of Educational Services of the school district reported that several revisions have occurred since the evaluation instrument was developed. First, the number of criteria and the number of sections have been revised. Second, descriptive statements of teacher behavior have been added to each criterion to describe levels of performance.

The Teacher Performance Evaluation Instrument used during the period of the study contains 25 criteria grouped in five sections: (a) Productive Teaching Techniques; (b) Organized, Structured Class Management; (c) Intellectual

Stimulation; (d) Positive Interpersonal Relations; and (e) Professional Job Responsibilities.

Section 1, Productive Teaching Techniques, is the largest, with 10 criteria. Section 1 criteria pertain to organizing instruction around objectives, using a variety of teaching techniques, evaluating student progress, and providing feedback.

Section 2, Organized, Structured Class Management, contains three criteria which include organizing students for effective instruction and implementing the instructional plan.

Section 3, Intellectual Stimulation, is the smallest section with two criteria which pertain to providing opportunities for successful learning experiences and helping students develop efficient learning skills.

Section 4, Positive Interpersonal Relations, contains six criteria such as promoting a positive self-concept in students and demonstrating effective relationships with the administration, fellow teachers, and parents.

Section 5, Professional Job Responsibilities, contains four criteria which pertain to professional growth and assuming responsibilities outside the classroom. See Appendix C for the Summative Report of the Teacher Evaluation Instrument.

Between the development of the instrument in 1973 and its use in this study, descriptive statements of teacher behavior were added to each criterion to describe levels of performance. This series of descriptive statements is known as a Behaviorally Anchored Rating Scale or BARS (Manatt, 1987). Manatt reported that this type of scale "increases the likelihood of scatter (namely, spreading out the ratings so that all do not receive a superior rating)." Four statements describe levels of performance for each criterion. A "Not Observed" choice is also available.

The lowest of the four performance levels describes the teacher's behavior as not exhibiting the desired characteristics. The next-to-lowest performance level describes the teacher's behavior as inconsistent in meeting the criteria. The next-to-highest performance level is labeled "Standard" and describes the teacher's behavior as meeting the standard in the criteria. The highest of the four levels describes the teacher's behavior as not only meeting the standard but exceeding it.

Below is an example of the four performance levels for the first criterion in the district's evaluation instrument:

Section 1: Productive Teaching Techniques

Criterion 1: Communicates effectively with students.

Performance Level	Descriptive Statements of Levels of Performance (Behaviorally Anchored Rating Scale)
(Lowest)*	Communications from the teacher are frequently unclear; students often appear confused.
(Next to lowest)*	Communications from the teacher are usually clear but student input is not encouraged.
Standard	Communications from the teacher are clear. Relevant dialogue is encouraged.
(Highest)*	In addition to meeting the standard, the teacher is extremely skillful in using a variety of verbal and nonverbal communications.

*Above terms in parentheses do not appear on the evaluation instrument.

The instrument was not changed during the three-year period under study. (See Appendix D for the Teacher Evaluation Instrument with Levels of Performance.)

The eight premises upon which the evaluation instrument was founded were compared to the four goals of a defensible evaluation process identified by Darling-Hammond et al. (1983). They described a defensible evaluation process as one that allows schools to balance the four goals of sorting teachers, maintaining staff morale, maintaining organizational integrity, and promoting incremental change. The school district's eight premises developed in 1973

appear to include the four goals identified by Darling-Hammond et al. The first goal, sorting teachers, was a part of the school district's premise two (singling out), premise four (differentiating levels of performance), and premise eight (providing information regarding retention and dismissal). The second goal identified by Darling-Hammond et al., maintaining staff morale, was included in premise two (identifying strengths), premise five (conducting evaluations openly), and premise six (respecting the uniqueness of the individual). Goal three, maintaining organizational integrity, was included in premise one (recognizing the right of the school community to expect valid and continuous evaluations of all district employees). Goal four, promoting incremental change, was encompassed in premise two (developing a course of action to improve upon deficiencies).

Data Collection Procedures

The researcher developed an identification system with school district personnel so that anonymity of teacher evaluation ratings would be assured. The identification system used code numbers rather than teacher names.

The researcher began with the district's list of secondary certificated staff members, their years of seniority, and subjects taught. The district also provided a second list of teacher names by building and the most

recent year each teacher was evaluated. The researcher identified 86 teachers on the seniority list who met the following three criteria:

1. had been evaluated one or more times during the three-year period, 1988-89, 1989-90, or 1990-91
2. taught at the secondary level (at the high school or one of the two junior highs)
3. taught one of the four core subjects (language arts, math, social studies, or science)

From information provided by the district, the researcher recorded each teacher's formal education as recognized by the district at the beginning of the year in which the teacher was evaluated and the gender of the teacher.

The researcher gave the data list to an uninterested intermediary who randomly assigned a code number to each teacher name. The intermediary forwarded the data list which included each teacher's name and the code number to the school district. The intermediary returned to the researcher the data list after replacing each teacher's name with the code number.

The school district then made a photocopy of each teacher's most recent evaluation, removed each teacher's name, added the code number, and forwarded the evaluation photocopies to the researcher. The school district returned

to the intermediary the list with the teacher names and code numbers. The intermediary retained the data list which assured that the identity of subjects was concealed from the investigator. The researcher added to the data list the 25 evaluation ratings by matching code numbers on the evaluations to those on the data list. The researcher assigned identifiers to the evaluators and recorded their gender.

The data set was comprised of the following information on the 79 teachers representing the population of this study:

1. Evaluation rating on each of the 25 criteria
2. Number of years of teaching experience recognized by the district
3. Amount of formal education recognized by the district
4. Teacher gender
5. Subject taught
6. Building assignment
7. Evaluator identifier
8. Evaluator gender
9. Year of most recent evaluation

Ratings on the evaluation instrument were translated to points as follows: the lowest rating, 1; second-lowest rating, 2; meets standard, 3; highest rating, 4. The

maximum number of points which could be received on the evaluation instrument was 100. (See Appendix E for the Codes Used in Data Collection.)

Number of years of teaching experience recognized by the district was recorded as the number of years of teaching experience completed prior to the year of evaluation. The negotiated contract between the district and the teachers' association defines the maximum number of years of experience the district can recognize when a teacher is hired. A teacher could have more years of experience than recognized by the district. For example, if a teacher with 10 years of experience were hired, the district would recognize a maximum of 8 years of experience based on the negotiated contract. The number of years of experience recognized by a district is one factor which determines teacher salary.

Formal education recognized by the district was recorded as earned degree(s) plus additional hours, e.g., Master's degree and 30 hours. Only earned degrees and credit hours recognized by the district were used. For purposes of the study, formal education was separated into two groups: group 1 was less than a Master's degree; group 2 was a Master's degree or more.

Data on teacher gender, subject taught, building assignment, evaluator, and year evaluated were collected

since these variables could offer alternative explanations for differences in evaluation ratings in addition to teacher experience and formal education.

Research Design

The purpose of this ex-post facto research study was to determine if two variables, experience and formal education, are associated with the criterion variable, evaluation ratings. Two sets of attributes were measured: (a) evaluation ratings and experience; and (b) evaluation ratings and formal education.

A contingency coefficient was used to measure the extent of association between the two sets of attributes. A contingency coefficient is used with nominally scaled variables (Siegel, 1956).

Research Hypotheses

This study was based on two hypotheses:

Hypothesis 1: Teacher evaluation ratings are not associated with teacher experience.

Hypothesis 2: Teacher evaluation ratings are not associated with teacher education.

In addition to looking at teacher performance globally using the total evaluation ratings, teacher performance on each section of the evaluation instrument was used to look

more specifically for possible relationships between ratings, experience, and formal education.

Within each of the two major hypotheses were five additional, or subset, hypotheses. The subset hypotheses used section ratings on the evaluation instrument instead of the total ratings on the evaluation instrument. The subset hypotheses pertained to the relationship between the section ratings and the experience and formal education of the subjects.

The subset hypotheses under Hypothesis 1 pertain to the relationship between section ratings and teachers' experience.

Hypothesis 1.1: Teacher evaluation ratings in Section 1, Productive Teaching Techniques, are not associated with teachers' experience.

Hypothesis 1.2: Teacher evaluation ratings in Section 2, Organized, Structured Class Management, are not associated with teachers' experience.

Hypothesis 1.3: Teacher evaluation ratings in Section 3, Intellectual Stimulation, are not associated with teachers' experience.

Hypothesis 1.4: Teacher evaluation ratings in Section 4, Positive Interpersonal Relations, are not associated with teachers' experience.

Hypothesis 1.5: Teacher evaluation ratings in Section 5, Professional Job Responsibilities, are not associated with teachers' experience.

The subset hypotheses under Hypothesis 2 pertain to the relationship between section ratings and teachers' formal education.

Hypothesis 2.1: Teacher evaluation ratings in Section 1, Productive Teaching Techniques, are not associated with teachers' formal education.

Hypothesis 2.2: Teacher evaluation ratings in Section 2, Organized, Structured Class Management, are not associated with teachers' formal education.

Hypothesis 2.3: Teacher evaluation ratings in Section 3, Intellectual Stimulation, are not associated with teachers' formal education.

Hypothesis 2.4: Teacher evaluation ratings in Section 4, Positive Interpersonal Relations, are not associated with teachers' formal education.

Hypothesis 2.5: Teacher evaluation ratings in Section 5, Professional Job Responsibilities, are not associated with teachers' formal education.

Treatment of the Data

First, descriptive data for the population of 79 teachers were reviewed, comparing mean total evaluation ratings with teachers' experience and formal education. Second, an item analysis of the evaluation instrument was conducted to determine if the items discriminate between high-performance and low-performance teachers. To conduct an item analysis, Borg and Gall (1983) recommended selecting the 27% of the subjects with the highest criterion score and the 27% with the lowest criterion score. Hopkins et al. (1990) also recommended using the high and low 27% and noted other researchers' experience (p. 169). The high-performance 27% and low-performance 27% of the total evaluation ratings were identified for the item analysis (Borg & Gall, 1983; Hopkins, Stanley, & Hopkins, 1990). The proportion of teachers rated as exceeding district

standards, the highest rating, was recorded item by item. Difficulty and discrimination values were computed.

Third, high-performance and low-performance groups were identified. The high-performance group was composed of teachers whose scores were in the highest 27% of the population. The low-performance group was composed of teachers whose scores were in the lowest 27% of the population. Fourth, the hypotheses were tested by using contingency coefficients to determine the association between teachers' evaluation ratings and the teachers' experience and formal education. Contingency coefficients were also used to determine the association between teachers' evaluation ratings by section of the evaluation instrument and teachers' experience and formal education.

Fifth, the data set was further explored by reviewing the population, as well as the high- and low-performance groups, by teacher gender, subject taught, building, and evaluator for alternative explanations of the variability in evaluation ratings.

Chapter 4 contains the descriptive data on the population of 79 teachers as well as the item analysis results. Contingency coefficients were obtained to reveal the degree of association between teachers' total evaluation ratings and their experience and formal education.

Contingency coefficients were also obtained to reveal the degree of association between teachers' section evaluation ratings and their experience and formal education. Chapter 4 also contains an analysis of additional variables which may explain variability in the evaluation ratings.

Chapter 4

ANALYSIS OF THE DATA

Chapter 4 begins with an overview of the total evaluation ratings, experience, and formal education of the 79 teachers in the population. Second, the item analysis of the evaluation instrument is described. Third, the procedure to divide the population into high-performance and low-performance groups is detailed. Fourth, contingency coefficient results are reported showing the relationship between teachers' evaluation ratings and their experience and formal education. The chapter concludes with an analysis of other variables such as the teachers' gender, subject taught, building assignment, and evaluator which may contribute to the differences in evaluation ratings.

Descriptive Data: Population

Among the population of 79 teachers, the highest total evaluation rating received was 100; and the lowest was 73. The population's mean total evaluation rating was 87.3. (Appendix F shows the Total Evaluation Ratings by Frequency and Percentage.)

During the three years of the study, 27 teachers were evaluated in 1988-89, 32 in 1989-90, and 20 in 1990-91. The mean total evaluation rating by year of evaluation increased from 85.7 in 1988-89 to 89.2 in 1990-91. Analysis of

variance showed no significant difference among the means. (See Appendix G for the Mean Evaluation Ratings by Year Evaluated.)

The mean experience of the 79 teachers in the population was 12.2 years. The mean formal education was 1.49, meaning that slightly less than one-half of the teachers in the population had less than a Master's degree. (Appendix H shows the Mean Evaluation Ratings, Experience and Formal Education.)

Item Analysis

An item analysis of the 25 criteria on the evaluation instrument was conducted. As recommended by Borg and Gall (1983) and Hopkins et al. (1990), the highest 27% and lowest 27% of the population were identified by total evaluation ratings. Difficulty and discrimination levels were determined for each item on the evaluation instrument. Including tied total evaluation scores, the high-performance group consisted of 22 teachers; the low-performance group consisted of 24 teachers.

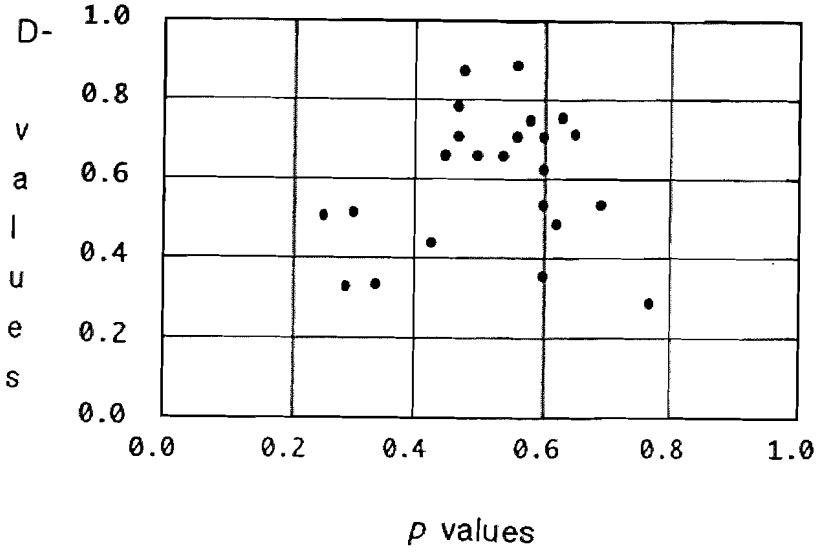
Item difficulty, or *p* value, was determined by adding item by item the percent of high-performance group members to the percent of low-performance group members rated as exceeding district standards and dividing by 2:

$$p = \frac{pH + pL}{2}.$$

According to Hopkins et al. (1990), the maximum potential measurement value of a test item is .5. Hopkins et al. reported that in the middle range of difficulty (25% to 75%) all items have the potential for very high discrimination. Very easy or very difficult items, however, fail to assess individual differences. The range of *p* values from the evaluation instrument was from .25 to .77. (See Table 1 for plotted *p* values and Appendix I for *P* Values by Item.)

Item discrimination was determined by subtracting the percentage of the low-performance group members rated as exceeding district standards from the percent of the high-performance group members rated as exceeding district standards: $D = pH - pL$. According to Hopkins et al. (1990), test items that yield a discrimination index of .35 or more are relatively high in discrimination. Test items with D-values below .2 are relatively low in discrimination (Hopkins et al., p. 270). Figure 1 shows the D-values of the items on the evaluation instrument. Applying the scale by Hopkins et al. to the items on the evaluation instrument, 22 of the 25 items discriminate between the high-performance and low-performance groups based on total evaluation ratings. (See Appendix I for the discrimination values by item.)

Figure 1. Item analysis of evaluation instrument based on *Exceeds* ratings: High- and low-performance groups.



Descriptive Data: High-performance and Low-performance Groups

From the population of 79 teachers, high- and low-performance groups were identified and their characteristics compared to determine if they differed in experience or formal education. The high- and low-performance groups were also compared to the population.

The high- and low-performance groups were formed by identifying the highest 27% and lowest 27% of the total evaluation ratings. Because using the highest and lowest 27% as comparative groups in the item analysis revealed a nearly model pattern of difficulty and discrimination

levels, the same procedure was used to identify two groups of teachers for further analysis. The high-performance group was composed of 22 teachers; the low-performance group was composed of 24 teachers. Difference in group size was due to duplicate scores in total evaluation ratings.

The range of total evaluation ratings for the high-performance group was from 93 to 100; the range for the low-performance group was from 73 to 83. The lowest rating in the high-performance group, 93, would result from two-thirds of the ratings being 4s (exceeds district standards) and one-third of the ratings being 3s (meets district standards). The highest rating in the low-performance group, 83, would result from two-thirds of the evaluation ratings being 3s and one-third being 4s. The lowest total evaluation rating for a teacher in the low-performance group, 73, represents a less-than-satisfactory performance and an average rating of 2.9 on the 25 items on the evaluation instrument. (See Appendix J for the Frequency of Evaluation Ratings for the High-performance Group and Appendix K for the Low-performance Group.)

The mean total evaluation rating for the high-performance group was 95.7; the mean total evaluation rating for the low-performance group was 80.0. (See Appendix L for the Mean Evaluation Ratings by Year Evaluated for the High- and Low-performance Groups.)

The teachers in the high-performance group had a mean years of experience (12.5 years) nearly identical to the mean years of experience of teachers in the low-performance group (13.5 years).

The mean formal education levels of teachers in the high-performance and low-performance groups differed. The mean formal education level of the high-performance group was notably higher (1.73) than the low-performance group (1.50). The mean formal education level of the low-performance group, however, was very close to the mean formal education level of the population (1.50 compared to 1.49), meaning that about one-half of the group members had less than a Master's degree. (See Appendix M for the Mean Evaluation Ratings, Experience, and Formal Education for the High- and Low-performance Groups.)

Hypotheses of the Study

This study poses two major hypotheses and 10 sub-hypotheses related to the degree of association between teachers' evaluation ratings and teachers' experience and formal education.

Hypothesis 1

The first hypothesis pertains to the relationship between total evaluation ratings and teacher experience:

Hypothesis 1: Teacher evaluation ratings are not associated with experience.

Since the literature contained examples of several different experience groupings, four different experience groupings were used.

First, the Teacher Incentive Programs in the Public Schools (Bobbitt, 1989) used six experience categories: (a) Less than 6 years, (b) 6 to 10, (c) 11 to 15, (d) 16 to 20, (e) 21 to 25, and (f) 26 years and above. These categories were identified as Experience Grouping 1.

Second, the Florida Performance Measurement System norming study used the following five categories: (a) 1 year or less, (b) 2 to 4 years, (c) 5 to 9 years, (d) 10 to 14 years, and (e) 15 years and above (Micceri, 1984). Due to the population distribution of this study, the first two categories were combined. These categories were identified as Experience Grouping 2.

Third, Schwab (1991) used the following four categories: (a) Less than 5 years, (b) 5 to 12 years, (c) 13 to 24 years, and (d) 25 years and above. These categories were identified as Experience Grouping 3.

Fourth, the researcher developed another experience grouping using these categories: (a) Less than 5 years, (b) 5 to 10 years, (c) 11 to 16 years, (d) 17 to 21 years, and (e) 22 years and above. This sequence separated

probationary teachers from non-probationary teachers and, considering a typical evaluation pattern in the district, tended to place non-probationary teachers into a new category after every other evaluation. Although non-probationary teachers were typically evaluated every third year, evaluations could occur more frequently. These categories were identified as Experience Grouping 4.

After grouping the 46 teachers into the high- and low-performance groups, contingency coefficients were run by each of the four experience groupings. The contingency coefficients were suspect, however, due to insufficient cell size. Siegel (1956) recommended combining categories if more than 20% of the cells have expected frequencies of less than 5.

A fifth experience grouping was developed by the researcher to ensure sufficient cell size and to use experience groupings similar to those in the literature. Experience Grouping 5 contained the following categories: (a) Less than 6 years, (b) 7 to 14 years, (c) 15 to 18 years, and (d) 19 years and above.

Table 1 shows the frequency distribution of teachers in the highest 27% and lowest 27% of total evaluation ratings grouped by years of experience. The upper limit for a contingency coefficient is a function of the number of categories. Since the contingency coefficient is calculated

from a 2 x 4 table, the upper limit is .87 (Siegel, 1956). The table yielded a contingency coefficient, *C*, of .20, which represents the degree of association between total evaluation scores and experience. The degree of association between total evaluation ratings and experience is weak. Hypothesis 1 is accepted.

Table 1

Experience: High- and Low-performance Groups

Years of Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	5	6	11
7 to 14 years	7	6	13
15 to 18 years	6	4	10
19 years and more	4	8	12
Total	22	24	46

Hypothesis 2

The second major hypothesis pertained to the relationship between teacher evaluation ratings and formal education:

Hypothesis 2: Teacher evaluation ratings are not associated with formal education.

Table 2 shows the frequency distribution of teachers in the highest 27% and lowest 27% of total evaluation ratings grouped by formal education. The upper limit for a 2 x 2 contingency table is .77. The contingency table yielded a contingency coefficient, *C*, of .23, which represents the relationship between total evaluation scores and formal education. The degree of association between total evaluation ratings and education is weak. Hypothesis 2 is accepted.

Table 2

Formal Education: High- and Low-performance Groups

Formal Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	6	12	18
Master's degree or more	16	12	28
Total	22	24	46

Evaluation Ratings by Section of Evaluation Instrument

In addition to using the total of all evaluation ratings on the evaluation instrument, the section evaluation ratings were used. Each section included clusters of

teacher behaviors such as using productive teaching techniques, organizing students for instruction, providing intellectual stimulation, demonstrating positive interpersonal relations, and fulfilling professional responsibilities. (See Appendix N for Evaluation Scores by Section for the Population.) The total of each section rating was used to determine if a relationship exists between teachers' section ratings and their experience and formal education.

The highest 27% and lowest 27% of each section's evaluation ratings were identified. (See Appendix O for the Evaluation Scores by Section for the High- and Low-Performance Groups.) The number of teachers in the high 27% and low 27% within each section varied due to duplicate scores. Without duplicate scores, the high-performance and low-performance groups in each section would have included 21 scores.

Section Ratings and Teacher Experience

Hypothesis 1.1: Evaluation ratings in Section 1, Productive Teaching Techniques, are not associated with teacher experience.

Table 3 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 1 ratings grouped by years of experience. The contingency table yielded a

contingency coefficient, *C*, of .16, which represents a weak relationship between Section 1 evaluation ratings and experience. Hypothesis 1.1 is accepted.

Table 3

Section 1 Evaluation Ratings by Experience

Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	6	9	15
7 to 14 years	8	7	15
15 to 18 years	6	4	10
19 years and more	5	7	12
Total	25	27	52

Hypothesis 1.2: Evaluation ratings in Section 2, Organized, Structured Class Management, are not associated with teacher experience.

Table 4 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 2 ratings grouped by years of experience. The contingency table yielded a contingency coefficient, *C*, of .31, which represents a moderate relationship between Section 2 evaluation ratings

and experience. Hypothesis 1.2 is accepted although the moderate relationship is noted.

Table 4

Section 2 Evaluation Ratings by Experience

Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	6	12	18
7 to 14 years	9	7	16
15 to 18 years	6	5	11
19 years and more	2	11	13
Total	23	35	58

Hypothesis 1.3: Teacher evaluation ratings in Section 3, Intellectual Stimulation, are not associated with teacher experience.

Table 5 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 3 ratings grouped by years of experience. The contingency table yielded a contingency coefficient, *C*, of .16, which represents a weak relationship between Section 3 evaluation ratings and experience. Hypothesis 1.3 is accepted.

Table 5

Section 3 Evaluation Ratings by Experience

Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	11	14	25
7 to 14 years	7	16	23
15 to 18 years	6	7	13
19 years and more	5	13	18
Total	29	50	79

Hypothesis 1.4: Evaluation ratings in Section 4, Positive Interpersonal Relations, are not associated with teacher experience.

Table 6 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 4 ratings grouped by years of experience. The contingency table yielded a contingency coefficient, *C*, of .14, which represents a weak relationship between Section 4 evaluation ratings and experience. Hypothesis 1.4 is accepted.

Table 6

Section 4 Evaluation Ratings by Experience

Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	10	10	20
7 to 14 years	9	12	21
15 to 18 years	6	5	11
19 years and more	4	8	12
Total	29	35	64

Hypothesis 1.5: Evaluation ratings in Section 5, Professional Job Responsibilities, are not associated with teacher experience.

Table 7 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 5 ratings grouped by years of experience. The contingency table yielded a contingency coefficient, *C*, of .18, which represents a weak relationship between Section 5 evaluation ratings and experience. Hypothesis 1.5 is accepted.

Table 7

Section 5 Evaluation Ratings by Experience

Experience	No. Teachers High-performance Group	Low-performance Group	Total
Less than 7 years	8	11	19
7 to 14 years	10	9	19
15 to 18 years	6	6	12
19 years and more	4	10	14
Total	28	36	64

Section Evaluation Ratings and Formal Education

Section evaluation ratings were also used to determine if a relationship exists between the five clusters of teacher behaviors and teachers' formal education. Using the high and low 27% of each section score, formal education was grouped by the same two categories: less than Master's degree and Master's degree or more.

Hypothesis 2.1: Teacher evaluation ratings in Section 1, Productive Teaching Techniques, are not associated with teacher formal education.

Table 8 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 1 ratings grouped

by formal education. The contingency table yielded a contingency coefficient, *C*, of .23, which represents a weak relationship between Section 1 evaluation ratings and formal education. Hypothesis 2.1 is accepted.

Table 8

Section 1 Evaluation Ratings by Formal Education

Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	8	15	23
Master's degree or more	1	12	29
Total	25	27	52

Hypothesis 2.2: Teacher evaluation ratings in Section 2, Organized, Structured Class Management, are not associated with teacher formal education.

Table 9 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 2 ratings grouped by formal education. The contingency table yielded a contingency coefficient, *C*, of .21, which represents a weak

relationship between Section 2 evaluation ratings and formal education. Hypothesis 2.2 is accepted.

Table 9

Section 2 Evaluation Ratings by Formal Education

Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	8	20	28
Master's degree or more	15	15	30
Total	23	35	58

Hypothesis 2.3: Teacher evaluation ratings in Section 3, Intellectual Stimulation, are not associated with teacher formal education.

Table 10 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 3 ratings grouped by formal education. The contingency table yielded a contingency coefficient, *C*, of .19, which represents a weak relationship between Section 3 evaluation ratings and formal education. Hypothesis 2.3 is accepted.

Table 10

Section 3 Evaluation Ratings by Formal Education

Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	11	29	40
Master's degree or more	18	21	39
Total	29	50	79

Hypothesis 2.4: Teacher evaluation ratings in Section 4, Positive Interpersonal Relations, are not associated with teacher formal education.

Table 11 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 4 ratings grouped by formal education. The contingency table yielded a contingency coefficient, *C*, of .13, which represents a weak relationship between Section 4 evaluation ratings and formal education. Hypothesis 2.4 is accepted.

Table 11

Section 4 Evaluation Ratings by Formal Education

Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	12	19	31
Master's degree or more	17	16	33
Total	29	35	64

Hypothesis 2.5: Teacher evaluation ratings in Section 5, Professional Job Responsibilities, are not associated with teacher formal education.

Table 12 shows the frequency distribution of teachers in the highest 27% and lowest 27% of Section 5 ratings grouped by formal education. The contingency table yielded a contingency coefficient, *C*, of .24, which represents a weak relationship between Section 5 evaluation ratings and formal education. Hypothesis 2.5 is accepted.

Table 12

Section 5 Evaluation Ratings by Formal Education

Education	No. Teachers High-performance Group	Low-performance Group	Total
Less than Master's degree	14	22	36
Master's degree or more	18	10	28
Total	32	32	64

Descriptive Data: Grouping by Other Variables

Population

The population was grouped by teacher gender, subject taught, building assignment, and evaluator to investigate alternative explanations for differences in evaluation ratings when compared to experience and formal education.

Variable: Gender

If evaluation ratings are associated with years of experience, one would expect to find higher mean evaluation ratings among males since males had notably more years of experience than females. (See Appendix P.) The contrary was true. Although the mean years of experience of males (14.5 years) was greater than that of females (9.7 years),

males' mean evaluation rating (86.5) was lower than the females' (88.2). Therefore, it does not appear that a higher mean years of experience is associated with higher mean evaluation ratings when the data are grouped by gender.

If formal education is associated with evaluation ratings, one would expect to find a higher mean rating for males since males had a mean formal education level of 1.56 compared to a mean formal education level of 1.42 for females. As noted in the above paragraph, however, males had a lower mean evaluation rating than females. Therefore, it does not appear that a higher formal education is associated with higher mean evaluation ratings when the data are grouped by gender.

Variable: Subject Taught

The distribution of language arts, math, social studies, and science teachers in the population varied little, with a high of 21 language arts teachers to a low of 18 science teachers. (See Appendix Q.) Grouping evaluation ratings, experience, and formal education by subject taught showed little difference among the four subject areas.

Variable: Building

The difference in the mean evaluation ratings across buildings was relatively small. (See Appendix R.) However, the building with the highest mean rating, junior high 2,

had a mean years of experience greater than one building and less than the other. The building with the least mean years of experience, junior high 1, had a mean evaluation rating slightly higher than the building with the most years of experience, the high school. Contrary to what might be expected, then, it does not appear that evaluation ratings are associated with years of experience when the data are viewed by building.

Although there was little variability among the mean evaluation ratings among the three buildings, it does not appear that a higher mean evaluation rating is associated with a higher level of formal education when the data are grouped by building.

Variable: Evaluator

The population of 79 teachers was evaluated by a total of seven evaluators, all of whom were male building or assistant building principals. (See Appendix S.) Four evaluators were from the high school, including a principal and three assistant principals. The two evaluators from junior high 1 were successive principals. The one evaluator from junior high 2 was the principal.

The highest mean evaluation rating by an evaluator was 90.2 by evaluator 1 at the high school. The lowest mean evaluation rating was 83.6 by evaluator 5 at junior high 1. The number of teachers evaluated ranged from a high of 24 by

evaluator 2 at the high school to a low of 5 teachers by evaluator 5 at junior high 1.

An analysis of variance comparing the mean evaluation ratings by evaluators showed an F value of 2.27 and a significance level of .046. The differences in mean evaluation ratings may be explained by the number of teachers evaluated.

A post hoc comparison was conducted to further investigate the relationship between differences in mean evaluation ratings by evaluator. Using the high and low mean evaluation ratings (90.2 by evaluator 1 and 83.6 by evaluator 5), the Scheffé post hoc test showed an F value of 2.99 which failed to meet the critical value of 13. It does not appear that differences in evaluation ratings can be explained by differences in evaluators.

If evaluation ratings are associated with experience, one would expect to find that the groups by evaluator with the highest and lowest mean total evaluation ratings would have the most and least years of experience. Evaluator Group 1 had the highest mean total evaluation ratings as well as the second-highest mean experience, 15.6 years. However, the lowest evaluator group, Evaluator Group 5, had the third-highest years of experience. The evaluator group with the lowest mean experience, Evaluator Group 6, had the second-highest mean total evaluation rating. It does not

appear a relationship exists between total evaluation ratings and experience when ratings are grouped by evaluator.

If evaluation ratings are related to formal education, one would expect to find that the evaluator groups with the highest mean total evaluation ratings would have the highest mean formal education levels. Most groups' mean formal education was near the population mean (1.49). However, Evaluator Group 1 had notably higher mean formal education (1.78) as well as the highest mean total evaluation rating, 90.2. The two evaluator groups with the least amount of formal education were Evaluator Groups 3 and 6. Evaluator Group 3 had mean total evaluation ratings very close to the population mean (87.2 compared to 87.3), and Evaluator Group 6 had the second-highest mean total evaluation rating, 89.7. It does not appear that the extent of teachers' formal education corresponds with their mean total evaluation ratings when teacher total evaluation ratings are grouped by evaluator. (See Appendix T for the rankings by evaluation ratings, experience, and formal education.) A consistent pattern of a high rank in teachers' ratings and teachers' experience or formal education is not apparent.

Descriptive Data: Grouping by Other Variables

High-performance and Low-performance Groups

Variable: Gender Among High- and Low-Performance Groups

If evaluation ratings are associated with experience, one would expect to find that high-performance teachers, both males and females, would have notably more experience than low-performance teachers. Among males, this was not true. High-performance males had less experience than low-performance males. (See Appendix U.) Females in the high-performance group, however, did have more experience than females in the low-performance group. Among the four groups, the low-performance males had the most experience, and the low-performance females had the least experience. It does not appear, therefore, that evaluation ratings are associated with experience among males. A relationship between total evaluation ratings and experience may be indicated among females.

If formal education is associated with evaluation ratings, one would expect to find a higher mean for education among males and females in the high group compared to males and females in the low group. Males in the high group, however, had less formal education than males in the low group, 1.55 compared to 1.67 respectively. Females in the high group, though, did have more education than females in the low group, 1.91 compared to 1.22, respectively.

Therefore, a relationship may exist between total evaluation ratings and formal education among females although such a relationship is not apparent among males.

Variable: Subject Taught

Several instances were apparent in which high-performance group teachers had more experience than low-performance group teachers. Language arts and math teachers in the high-performance groups had more experience than their counterparts in the low-performance groups. The high science group showed an opposite pattern. High-performance group science teachers had low experience (7.0 years) while low-performance group science teachers had high experience (18.4 years).

If formal education is associated with evaluation ratings, one would expect to find a higher mean for education among high-performance group teachers than among low-performance group teachers when separated by subject taught. This was the case for three of the four subject areas: language arts (2.00 and 1.38), math (1.60 and 1.29), social studies (1.71 and 1.50). A different pattern was evident among science teachers. High-performance group science teachers had a mean education of 1.67 compared to low-performance group science teachers who had a mean education of 2.00. (See Appendix V.) All five low-performance group science teachers had at least a

Master's degree while four of the six high-performance group science teachers had a Master's degree.

Variable: Building

If experience is associated with evaluation ratings, one would expect to find that high-performance group teachers have substantially more experience than low-performance group teachers within the same building. (See Appendix W.) No obvious pattern between evaluation ratings and experience was apparent using the six groups. Group sizes varied, however, from 2 to 18.

If evaluation ratings are associated with formal education, one would expect to find that high-performance group teachers have substantially more formal education than low-performance group teachers within the same building. In each of the three buildings, the high-performance group had more formal education than the low-performance group.

Variable: Evaluator

Data showing the evaluation ratings, experience, and formal education of high-performance and low-performance groups by evaluator were reviewed. Due to the small size of the groups, the data are not included. (See Appendix X, Number of Cases and Mean Ratings by Evaluator: High- and Low-performance Groups.)

Summary of Chapter 4

An item analysis was conducted to determine which items discriminated between the high- and low-performance groups. The item analysis showed that nearly all of the items on the evaluation instrument met desirable levels of difficulty and discrimination (Hopkins et al., 1990).

Only a weak relationship was found between teacher evaluation ratings and experience. A similarly weak relationship was found between teacher evaluation ratings and formal education. Contingency coefficients revealed only weak relationships between section evaluation ratings and teachers' experience and formal education. The strongest relationship between section ratings and teacher experience was in Organized, Structured Class Management, Section 2, of the evaluation instrument. The *C* of .31 shows a moderate relationship. Table 13 summarizes the relationship between section ratings and teacher experience.

Table 13

Relationship between Section Evaluation Ratings and Experience

Section	Contingency Coefficient	Upper Limit	Relationship Between Section Ratings & Experience
1	.16	.87	Weak
2	.31	.87	Moderate
3	.16	.87	Weak
4	.14	.87	Weak
5	.18	.87	Weak

Table 14 summarizes the relationship between section evaluation ratings and formal education.

Table 14

Relationship between Section Evaluation Ratings and Formal Education

Section	Contingency Coefficient	Upper Limit	Relationship Between Section Ratings & Experience
1	.23	.71	Weak
2	.21	.71	Weak
3	.19	.71	Weak
4	.13	.71	Weak
5	.24	.71	Weak

Chapter 4 concludes with descriptive data for both the population and the high- and low-performance groups. Chapter 5 contains an overview of the study, including its purpose, procedures, results, and conclusions. Suggestions for further study are included.

Chapter 5

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Summary of the Study

The purpose of Chapter 5 is to summarize the study and to interpret the findings presented in Chapter 4. This study is one response to the many calls for school reform. A frequent recommendation in the school reform movement is that teachers' performance be linked to their salaries. Presently, teachers' experience and formal education are the criteria used to determine salaries in the teaching profession.

The purpose of the study was to determine if teachers' performance is related to their experience and formal education. An item analysis of the 25 criteria on the evaluation instrument was conducted and revealed that nearly all items could discriminate at appropriate difficulty levels. Contingency coefficients were obtained which revealed only weak relationships between teachers' evaluation ratings and their experience and formal education when grouped by high- and low-performance. Contingency coefficients were also obtained to determine the relationships between each section's ratings and teachers' experience and formal education. Only a weak relationship

was found between section evaluation ratings and teachers' experience and formal education. A moderate relationship was indicated between section evaluation ratings in Section 2 and teachers' experience. Section 2 includes teacher behaviors in organizing and structuring class management.

The weak-to-moderate contingency coefficients are consistent with the reported results in the literature. Researchers (Bastress, 1980; Borich, 1977; Epstein, 1985; Hoogeveen & Gutkin, 1986; Micceri, 1984) reported no relationship between years of teaching experience and evaluation ratings. Borich (1977) reported that experience variables have been "almost worthless in predicting teacher performance" but questioned "if the variables were measured too grossly to yield significant findings" (p. 14). This study attempted to measure more specifically the variable of teacher performance by limiting the subjects to secondary level teachers in four subject areas and by using section ratings in addition to the total ratings.

The weak contingency coefficients showing the relationship between evaluation ratings and formal education are consistent with the literature. Researchers (Bastress, 1980; Epstein, 1985; Micceri, 1984) found no relationship between evaluation ratings and formal education.

Conclusions

Results of this study show that: (a) there is no relationship between teachers' evaluation ratings and their experience; (b) there is no relationship between teachers' evaluation ratings and their formal education; (c) there is no relationship between section evaluation ratings and experience with the exception that a moderate relationship was indicated between Section 2, Organized, Structured Class Management and teachers' experience; (d) there is no relationship between section evaluation ratings and formal education.

When teachers were separated into high- and low-performance groups, the two middle experience categories (7 to 14 years and 15 to 18 years) had more teachers in the high-performance group than in the low-performance group. (See Table 1.) Teachers in the lowest and highest experience categories (less than 7 years of experience and more than 18 years of experience) were more likely to be in the low-performance group than in the high-performance group. Among the 11 teachers with less than 7 years of experience, 6 teachers were in the low-performance group and 5 teachers were in the high-performance group. Of the 12 teachers with 19 and more years of experience, 8 were in the low-performance group and 4 were in the high-performance group. The teachers in the highest category of years of

experience were more likely to be in the low-performance group than teachers in any of the other three experience categories. A question for future study is if experience is positively associated with evaluation ratings in the middle-experience years while negatively associated in the lower and higher years of experience; in other words, if a non-linear relationship exists.

Among the 28 teachers with Master's degrees, 16 fell into the high-performance group while 12 fell into the low-performance group. (See Table 2.) Among the 18 teachers with less than a Master's degree, only 6 fell into the high-performance group while 12 fell into the low-performance group. Among the 22 teachers in the high-performance group, 16 (73%) had Master's degrees while among the 24 teachers in the low-performance group, only 12 (50%) had Master's degrees. Although the contingency coefficient was not high enough to support an association between evaluation ratings and formal education, it does appear that high-performance teachers are more likely to have a Master's degree than not. It appears that low-performance teachers are equally as likely to have a Master's degree as not to have one.

The low-performance science group is worth noting in its contrast to other subject-taught groups and to the high-performance science group. The low-performance science

group had more formal education (2.0) than any other group, with the exception of the language arts teachers with whom they tied. In addition, the low-performance science group had the highest mean years of experience (18.4 years) among all 8 groups and were notably higher than the high-performance science group in experience (18.4 years compared to 7.0 years). The high years of experience and high education of the low-performance science group contributed to the lack of a relationship between teachers' evaluation ratings and experience and formal education.

Although not directly related to the hypotheses of this study, mean evaluation ratings of the population of 79 teachers yielded findings of interest. Micceri (1984) noted a trend "although statistically non-significant and of small magnitude" (p. 24) in a steady drop in mean evaluation scores as experience increased. Although the mean ratings in this study which used different years of experience groups did not show a steady drop, the lowest mean ratings were found in the group with the most years of experience, 19 years and over. (See Table 15.)

Table 15

Mean Evaluation Ratings by Experience Group: Population

Experience	No. Cases	Evaluation Ratings
Less than 7 years	25	
Mean		87.2
S.D.		5.8
Range		77-100
7 to 14 years	23	
Mean		88.1
S.D.		6.4
Range		75-99
15 to 18 years	13	
Mean		88.4
S.D.		8.1
Range		73-98
19 years and more	18	
Mean		85.7
S.D.		6.5
Range		77-98
Population	79	
Mean		7.3
S.D.		6.5
Range		73-100

Among the population of 79 teachers, 40 teachers had less than a Master's degree and 39 had a Master's degree or more as recognized by the school district. The mean rating for the teachers with less than a Master's degree was somewhat lower than the mean rating for the teachers with a Master's degree or more. (See Table 16.)

Table 16

Comparison of Total Evaluation Ratings by Formal Education: Population

Formal Education	No. Cases	Evaluation Ratings
Less than Master's degree	40	
Mean		86.2
S.D.		5.4
Range		75-97
Master's degree or more	39	
Mean		88.5
S.D.		7.4
Range		73-100
Population	79	
Mean		87.3
S.D.		6.5
Range		73-100

Overall, the data show that a relationship exists neither between evaluation ratings and teachers' experience nor between evaluation ratings and formal education. Therefore, the reward system using experience and formal education as the criteria to determine teacher salaries is inconsistent with rewarding teachers based on their performance.

Recommendations

The results of this study can be used in making policy decisions about restructuring the teaching profession. Although the present reward system was instituted to improve the profession and to provide an incentive for professional development, it fails to reward performance. The strengths and weaknesses of the present reward system need to be assessed. Alternative reward systems need to be sought that will encourage and enable the teaching profession to develop the most capable professional staff possible.

Several avenues for future research appear especially noteworthy. First, are the highest-experience teachers those with relatively low evaluation ratings? If so, what changes can be made in their role, salary structure, or place in the organization so that the high-experience teachers can perform more effectively?

Further research could better evaluate the value of teaching experience by using both years of teaching experience recognized by the district and actual years of teaching experience. Although Bastress (1980) used both recognized years of teaching experience and actual years of teaching experience, his study was limited to student evaluations of secondary English teachers.

Further research could better evaluate the value of formal education by establishing criteria to more clearly define Master's degrees, as well as graduate hours, to make valid comparisons.

Further research could include the teacher's age as another variable. Age may play a more important role than was apparent in this study. Following teachers through their teaching careers within one district would provide a developmental picture of teaching skills in relation to years of experience and formal education, as well as other staff development training.

Continuing to conduct similar studies with various evaluation instruments, grade levels, and a variety of research designs will provide a more complete and valuable picture about the relationship between teaching performance and experience and formal education.

REFERENCES

Ashton, P., Crocker, L., & Olejnik, S. (1986). Does teacher education make a difference? A literature review and planning study. Tallahassee, FL: Institute on Student Assessment and Evaluation. Florida Department of Education. As quoted in Knapp, J., McNergney, R., Herbert, J., & York, H. (1990). Should a master's degree be required of all teachers? Journal of Teacher Education, 41(2), 27-37.

Barro, S. M., & Suter, L. (1988, July). International comparisons of teachers' salaries: An exploratory study. (Report No. CS 88-415). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Bastress, R. L. (1980). A study of the relationships of teacher experience, teacher-student relations, and teacher leadership style with teacher effectiveness in selected secondary school English classrooms. (Doctoral dissertation, University of Maryland, 1980). Dissertation Abstracts International, No. 8103862.

Bobbitt, S. A. (1989). Teacher incentive programs in the public schools. (Report No. CS 89-063). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Borg, W. R., & Gall, M. D. (1983). Educational research. New York: Longman.

Borich, G. D. (1977). The appraisal of teaching: Concepts and process. Reading, MA: Addison-Wesley Publishing.

Boyer, E. L. (1983). Foreword. In C. M. Feistritzer (Author). The condition of teaching - A state by state analysis. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Castetter, W. B. (1981). The personnel function in educational administration. New York: Macmillan.

Center for Public Sector Labor Relations. (1985, Spring). Teacher compensation and evaluation in public education. Dialogue: A review of labor-management cooperation in public education. Bloomington, IN: Author. (ERIC No. ED 259 475)

Cornett, L. (1989, April). Career ladder clearinghouse. Atlanta, GA: Southern Regional Education Board.

Darling-Hammond, L., Wise, A., & Pease, S. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53, 285-328.

Educational Excellence Act. (1987). Iowa Code, Chapter 294A.

Educational Research Service. (1988). Teacher evaluation: Practices and procedures (Report No. 219-21764). Arlington, VA: Author.

Epstein, J. L. (1985). A question of merit: Principals' and parents' evaluations of teachers. Educational Researcher, 14(7), 3-10.

Excellence in Education Task Force. (1984). First in the nation in education. Des Moines, IA: Iowa Department of Education.

Guernsey, M. A. (1986). Review of related literature and research: History of merit pay, differentiated staffing, and incentive pay programs. Washington, DC: National Institute of Education, U.S. Department of Education. (ERIC Document Reproduction Service No. ED 267 513).

Hoogeveen, K. & Gutkin, T. B. (1986). Collegial ratings among school personnel: An empirical examination of the merit pay concept. American Educational Research Journal, 23, 375-381.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). Educational and psychological measurement and evaluation (7th ed.). Englewood Cliffs, NJ: Simon & Schuster.

Iowa Association of School Boards. (1987). Strategies for excellence. Des Moines, IA: Iowa Association of School Boards.

Iowa Association of School Boards. (1989, May 10). Update, p. 1.

Iowa Department of Economic Development. (1991). Statistical profile of Iowa. Des Moines, IA: State of Iowa.

Knapp, J., McNergney, R., Herbert, J., & York, H. (1990). Should a master's degree be required of all teachers? Journal of Teacher Education, 41(2), 27-37.

Knienapfel, B. (1984). Merit pay for school administrators: A procedural guide. Arlington, VA: Educational Research Service.

Lepley, W. L. (1989, September 24). (Speaker). Speech to the city superintendents of Iowa. Newton, IA.

Lortie, D. C. (1975). School-teacher. Chicago, IL: The University of Chicago Press.

Manatt, R. P. (1987). Lessons from a comprehensive performance appraisal project. Educational Leadership, 44(7), 8-14.

McGreal, T. L. (1983). Successful teacher evaluation. Alexandria, VA: Association for Supervision and Curriculum Development.

Micceri, T. (1984). A technical report on the Florida Performance Measurement System norming study. Unpublished technical report. Tampa, FL: University of South Florida.

National Commission on Excellence in Education, The. (1983). A nation at risk: The imperative for educational reform. (GPO Publication No. 065-000-00177-2). Washington, DC: U.S. Government Printing Office.

National Education Association (NEA). (1985). Structuring teacher pay: Panaceas, problems and potential. West Haven, CT: Author.

Needham, N. R. (1982). Descheduling teacher salaries. Today's Education, 20(3), 103-107.

Northwest Regional Educational Laboratory. (1987). Summary of national reform recommendations. Portland, OR: Author.

Report on the development of a performance evaluation system. (June, 1974). Ames, IA: Iowa State University.

Robinson, G. E. (1983). Paying teachers for performance and productivity: Learning from experience. Arlington, VA: Educational Research Service.

Schwab, R. L. (1991). Can performance based salary programs motivate teachers? Insights from a case study. (Available from R. L. Schwab, Drake University, School of Education, 3206 University, Des Moines, IA 50311)

Shanker, A. (Speaker) (1988). Organizing schools for change based on the teacher's role. (Cassette Recording No. 612-89113C33). Alexandria, VA: Association for Supervision and Curriculum Development.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Tecker, G. H. (1985) Merit, measurement, and money. Alexandria, VA., National School Boards Association.

Travers, R. M. W. (1981). Criteria of good teaching. In Jason Millman (Ed.), Handbook of teacher evaluation (pp. 14-22). Beverly Hills, CA: Sage.

Appendix A

PREMISES FOR AN EVALUATION SYSTEM FOR TEACHERS

1. A school community has the right to expect that the school board, administrators, and the faculty of their district will conduct a valid and continuous evaluation of the service of all district employees.
2. The essential purpose of evaluation is to improve performance by identifying an educator's strengths and weaknesses and to provide guides to develop a course of action to improve upon deficiencies by
   a. singling out and strengthening the outstanding areas of teacher performance
   b. locating the deficiencies in the areas of teacher performance and designing methods by which to correct these deficiencies
   3. The professional educator desires improvement of his performance.
   4. It is possible to differentiate levels of performance.
   5. All evaluations of the educator's activities should be conducted openly and with the educator's full knowledge and awareness.

6. There is no single "model" educator that results in effective learning; we respect the uniqueness of each individual within the framework of the school.
7. Multiple sources of evaluation observation increase the validity of an evaluation report.
8. An evaluation system is an excellent basis for providing the best available information for decisions regarding retention and dismissal of personnel and providing information to be used as a basis for salary increases (movement on a salary schedule).

From: A Report on the Development of a Performance Evaluation System, June 1974, pp. 12-13.

Appendix B

CRITERIA AND PHILOSOPHIC PREMISES

The evaluation committee developed performance evaluation premises to serve as guidelines for the evaluation process for teachers, media specialists, and counselors. Subsequent committees have revised and refined and criteria.

Teachers

The role of the teacher is to facilitate all aspects of the student' growth and to help him/her reach his/her maximum potential. To this aim, the teacher job premises for this district are:

1. The teacher assesses the needs of students as individuals and as a group.
2. The teacher plans a program that meets the individual and group needs, interests, and abilities.
3. The teacher creates a classroom environment that is conducive to learning and appropriate to the maturity and interest of students.
4. The teacher guides the learning process toward the achievement of curriculum goals by establishing clear objectives and communicating these objectives to students.

5. The teacher employs a variety of instructional techniques and media to meet the students' needs.
6. The teacher evaluates pupils in terms of their academic, social, and emotional growth.
7. The teacher diagnoses the needs of, and provides instructional programs for, exceptional students and seeks the assistance of specialists as needed.
8. The teacher communicates effectively with colleagues, students, and parents.
9. The teacher assists in upholding the enforcing school rules, administrative regulations, and board policy.
10. The teacher promotes the development of student self-discipline and responsibility.
11. The teacher plans and supervises purposeful assignments for auxiliary personnel, e.g., aides, associates, and volunteers.
12. The teacher engages in appropriate studies and activities to improve professional competence.
13. The teacher contributes to the solution of educational problems, e.g., by serving on professional commissions.
14. The teacher carries out routine duties promptly and accurately.

15. The teacher assumes duties and responsibilities as may be assigned.
16. The teacher assists in selecting and updating books, equipment, and other instructional materials.
17. The teacher establishes and maintains cooperative relations with the community.

From: A Report on the Development of a Performance Evaluation System, June, 1974, pp. A-12.

Appendix C

SUMMATIVE REPORT OF THE TEACHER EVALUATION INSTRUMENT

TEACHER EVALUATION INSTRUMENT
(SUMMATIVE REPORT)

(Teacher's Name) _____ (Years Experience in District) _____ (Building) _____

(Evaluatee's Signature) _____ (Date) _____ (Evaluator's Signature) _____ (Date) _____

On the right hand side of this page, please check (✓) the level of performance which best describes the appraisee's performance on that item.

	LEVELS OF PERFORMANCE				
PRODUCTIVE TEACHING TECHNIQUES	N/D				STD
The teacher . . .					
1. Communicates effectively with students					
2. Motivates students					
3. Organizes instruction around objectives					
4. Demonstrates ability to select appropriate learning content					
5. Uses appropriate and available learning materials and resources					
6. Demonstrates ability to utilize a variety of appropriate teaching techniques					
7. Provides opportunities for individual differences					
8. Prepares appropriate evaluation activities					
9. Provides students with specific oral evaluative feedback					
10. Provides students with specific written evaluative feedback					
ORGANIZED, STRUCTURED CLASS MANAGEMENT					
The teacher . . .					
11. Organizes the educational setting					
12. Organizes and manages students for effective instruction					
13. Demonstrates skill in implementing instructional plan					
INTELLECTUAL STIMULATION					
The teacher . . .					
14. Provides opportunities for succesaful learning experiences for each pupil					
15. Helps students develop efficient learning skills and work habits					
POSITIVE INTERPERSONAL RELATIONS					
The teacher . . .					
16. Demonstrates sensitivity in relating to students					
17. Promotes positive self-concept					
18. Promotes self-discipline and responsibility					
19. Demonstrates effective interpersonal relationships with the administration					
20. Demonstrates effective interpersonal relationships with fellow teachers					
21. Cooperates and maintains an effective relationship with parents					
PROFESSIONAL JOB RESPONSIBILITIES					
The teacher . . .					
22. Shows professional growth					
23. Demonstrates evidence of professional ethics through support of school regulations and policies					
24. Assumes responsiblities outside the classroom as they relate to school					
25. Demonstrates willingness to keep curriculum and instructional practices current					

COMMENTS _____

Appendix D

TEACHER EVALUATION INSTRUMENT WITH LEVELS OF PERFORMANCE

TEACHER PERFORMANCE EVALUATION INSTRUMENT
(LEVELS OF PERFORMANCE)

PRODUCTIVE TEACHING TECHNIQUES

CRITERIA	LEVELS OF PERFORMANCE				
				STANDARD	
The teacher . . .					
1. Communicates effectively with students.	Not Observed	Communications from the teacher are frequently unclear; students often appear confused.	Communications from the teacher are usually clear but student input is not encouraged.	Communications from the teacher are clear. Relevant dialogue is encouraged.	In addition, to meeting the standard, the teacher is extremely skillful in using a variety of verbal and nonverbal communications.
2. Motivates students.	Not Observed	The teacher's unrealistic expectations dissuade students from performing according to their abilities.	The teacher usually motivates students to perform assigned tasks, but inconsistantly requires students to perform according to their abilities.	The teacher clearly expects and motivates students to perform assigned tasks according to their abilities.	In addition to meeting the standard, the teacher motivates students to achieve beyond previous performance levels.
3. Organizes instruction around objectives.	Not Observed	Instruction does not relate to the stated objectives.	Instruction marginally relates to the stated objectives.	The teacher clearly organizes instruction around the stated objectives.	In addition to meeting the standard, objectives are appropriate, specifically stated and measured.
4. Demonstrates ability to select appropriate learning content.	Not Observed	Learning content does not relate to approved curriculum guide(s).	Learning content is marginally related to the approved curriculum.	Learning content is related to the approved curriculum guide(s).	In addition to meeting the standard, the teacher shows initiative and leadership in review and development of curriculum.
5. Uses appropriate and available learning materials and resources.	Not Observed	The teacher does not use available resources and materials to aid instruction.	The teacher intermittently uses materials which are marginally related to lesson objectives.	The teacher uses available resources and materials which are appropriate to lesson objectives.	In addition to meeting the standard, the teacher seeks out and/or develops a variety of unique or original materials in the planned lesson.

TEACHER PERFORMANCE EVALUATION INSTRUMENT
(LEVELS OF PERFORMANCE)

PRODUCTIVE TEACHING TECHNIQUES (cont.)

				STANDARD	
6. Demonstrates ability to utilize a variety of appropriate teaching techniques.	Not Observed	The teacher shows little or no evidence of variety or appropriateness in teaching techniques.	The teacher intermittently uses variety in teaching techniques which are appropriate.	The teacher uses a variety of teaching techniques appropriate to the objectives of the lesson.	In addition to meeting the standard, the teacher develops unique teaching techniques to meet individual student needs.
7. Provides opportunities for individual differences.	Not Observed	The teacher does not provide for individual rates of learning and student capabilities.	The teacher intermittently provides for individual rates of learning and student capabilities.	The teacher provides for individual rates of learning and student capabilities.	In addition to meeting the standard, the teacher diagnoses individual learning styles and provides instruction accordingly.
8. Prepares appropriate evaluation activities.	Not Observed	The teacher uses evaluation activities which are irrelevant to the instructional objectives.	The teacher uses evaluation activities which are marginally related to the instructional objectives.	The teacher uses evaluation activities which are related to the instructional objectives.	In addition to meeting the standard, the teacher prepares a variety of evaluation activities which meet the needs of the individual students.
9. Provides students with specific oral evaluative feedback.	Not Observed	The teacher gives no evaluative feedback.	The teacher is inconsistent in giving feedback.	The teacher gives specific evaluative feedback.	In addition to meeting the standard, the teacher gives feedback with reinforcement and encouragment.
10. Provides students with specific written evaluative feedback.	Not Observed	The teacher gives no evaluative feedback.	The teacher is inconsistent in giving evaluative feedback.	The teacher gives specific evaluative feedback.	In addition to meeting the standard, the teacher gives feedback with reinforcement and encouragment.
11. Organizes the educational setting.	Not Observed	The teacher is disorganized in lesson preparation and organization.	The teacher intermittently presents materials in an organized manner.	Appropriate lesson preparation and organization of work is evident, i.e., materials are available; presentations progress logically.	In addition to meeting the standard, the teacher assesses and adjusts the educational setting to provide for a variety of learning styles.

TEACHER PERFORMANCE EVALUATION INSTRUMENT
(LEVELS OF PERFORMANCE)

ORGANIZED, STRUCTURED CLASS MANAGEMENT

CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . .					
12. Organizes and manages students for effective instruction.	Not Observed	The teacher displays no classroom control.	The teacher is inconsistent in maintaining classroom routine.	The teacher establishes, communicates and enforces the expected standards of behavior.	In addition to meeting the standard, the teacher uses techniques to help students demonstrate responsibility in a variety of work settings.
13. Demonstrates skill in implementing instructional plan.	Not Observed	The teacher shows little or no skill in implementing the instructional plan.	The teacher intermittently shows skill in implementing the instructional plan.	As the instructional plan is implemented, the teacher shows evidence of preparation and organization of the work with the objectives clearly in mind.	In addition to meeting the standard, the teacher uses a wide variety of techniques to implement the instructional plan.

INTELLECTUAL STIMULATION

CRITERIA	LEVELS OF PERFORMANCE				
				STANDARD	
The teacher . . .					
4. Provides opportunities for successful learning experiences for each pupil.	Not Observed	The teacher does not recognize individual student needs.	The teacher shows some concern for student needs.	The teacher recognizes and provides for student needs.	In addition to meeting the standard, the teacher shows sensitivity in helping the class to understand and provides for individual needs.
5. Helps students develop efficient learning skills and work habits.	Not Observed	The teacher makes no effort to help students develop efficient learning skills and work habits.	The teacher makes an effort to help students develop efficient learning skills and work habits but occasionally models them incorrectly.	The teacher is able to model and reinforce efficient learning skills and work habits.	In addition to meeting the standard, the teacher stimulates students to assume responsibility in a wide variety of settings.

TEACHER PERFORMANCE EVALUATION INSTRUMENT
(LEVELS OF PERFORMANCE)

POSITIVE INTERPERSONAL RELATIONS

CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . .					
16. Demonstrates sensitivity in relating to students.	Not Observed	The teacher is unresponsive to the needs of students.	The teacher intermittently shows sensitivity in communicating with students.	The teacher demonstrates sensitivity in relating to students.	In addition to meeting the standard, the teacher willingly provides extra time for individuals.
17. Promotes positive self-concept.	Not Observed	The teacher damages student self-concept by using excessive negative responses.	The teacher shows inconsistency in developing positive self-image for students; does not provide opportunities for success.	The teacher promotes positive self-image by providing opportunities for success for students.	In addition to meeting the standard, the teacher provides opportunities for students to achieve recognition for constructive behavior.
18. Promotes self-discipline and responsibility.	Not Observed	The teacher dissuades students from being responsible and self-disciplined through constant exposure to activities requiring dependency.	The teacher intermittently provides opportunities for students to develop responsibility and self-discipline.	The teacher provides opportunities for students to demonstrate responsible behavior.	In addition to meeting the standard, the teacher encourages students to demonstrate responsible behaviors in a wide variety of settings; i.e., through independent study, enrichment activities and group leadership roles.
19. Demonstrates effective interpersonal relationships with the administration.	Not Observed	The teacher is uncooperative in interactions with the administration.	The teacher is intermittently cooperative in relating to the administration.	The teacher has a good working relationship with the administration.	In addition to meeting the standard, the teacher works consistently with the administration to improve the educational program.
20. Demonstrates effective interpersonal relationships with fellow teachers.	Not Observed	The teacher is uncooperative with other teachers.	The teacher intermittently cooperates with other teachers.	The teacher cooperates by participating, contributing, and sharing.	In addition to meeting the standard, the teacher takes the lead in encouraging cooperation among teachers.

TEACHER PERFORMANCE EVALUATION INSTRUMENT
(LEVELS OF PERFORMANCE)

POSITIVE INTERPERSONAL RELATIONS (cont.)					
CRITERIA	LEVELS OF PERFORMANCE				
				STANDARD	
21. Cooperates and maintains an effective relationship with parents.	Not Observed	The teacher avoids interactions with parents.	The teacher intermittently cooperates and interacts with parents.	The teacher cooperates and effectively interacts with parents in implementing the child's educational program.	In addition to meeting the standard, the teacher seeks parent participation in the child's education.

PROFESSIONAL JOB RESPONSIBILITIES					
CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . .				STANDARD	
22. Shows professional growth.	Not Observed	The teacher shows no interest in professional growth activities.	When directed, the teacher attends professional growth activities.	The teacher seeks out and voluntarily participates in relevant growth activities.	In addition to meeting the standard, the teacher shares information with other staff members in regard to professional growth activities.
23. Demonstrates evidence of professional ethics through support of school regulations and policies.	Not Observed	The teacher openly refuses to comply with school regulations and policies.	The teacher intermittently cooperates in supporting school regulations and policies.	The teacher supports school regulations and policies.	In addition to meeting the standard, the teacher stays informed and/or participates in development and review of school regulations and policies.
24. Assumes responsibilities outside the classroom as they relate to school.	Not Observed	The teacher refuses to get involved in out-of-class responsibilities.	The teacher gets involved in out-of-class responsibilties only when directed.	The teacher assumes responsibility for the activities necessary for the smooth operation of the school.	In addition to meeting the standard, the teacher seeks and assumes "extra" responsibilities.
25. Demonstrates willingness to keep curriculum and instructional practices current.	Not Observed	The teacher uses curriculum and/or instructional practices that are outdated.	The teacher is reluctant to accept new materials and/or practices.	The teacher uses current materials and instructional practices.	In addition to meeting the standard, the teacher participates in curriculum review, development and improvement of instructional practices.

Appendix E

CODES USED IN DATA COLLECTION

Teacher #	randomly assigned by uninterested intermediary	
Year evaluated	1	1988-89
	2	1989-90
	3	1990-91
Education	1	less than a Master's degree
	2	Master's degree or more
Experience	number of years of experience recognized by the district	
Teacher gender	1	male
	2	female
Subject taught	1	language arts
	2	math
	3	social studies
	4	science
Building	1	high school
	2	junior high 1
	3	junior high 2
Evaluator	1	high school
	2	high school
	3	high school
	4	high school
	5	junior high 1
	6	junior high 1
	7	junior high 2
Evaluator gender	1	male (no evaluators were female)
Ratings·	4	exceeds district standards
	3	meets district standards
	2	needs improvement
	1	unsatisfactory
	0	not observed

Appendix F

TOTAL EVALUATION RATINGS BY FREQUENCY AND PERCENTAGE: POPULATION

Total Evaluation Ratings	Frequency	% of Population
100	1	1.3
99	1	1.3
98	2	2.5
97	4	5.1
96	4	5.1
95	2	2.5
94	4	5.1
93	4	5.1
92	1	1.3
91	1	1.3
90	6	7.6
89	3	3.8
88	2	2.5
87	3	3.8
86	7	8.8
85	5	6.3
84	5	6.3
83	5	6.3
82	5	6.3
81	2	2.5
80	3	3.8
79	3	3.8
78	1	1.3
77	2	2.5
75	2	2.5
73	1	1.3
Total	79	100.0

Appendix G

MEAN EVALUATION RATINGS BY YEAR EVALUATED:

POPULATION

Year	Number of Evaluations	% Total Evaluations	Mean Evaluation Ratings
1988-89	27	34%	85.7
1989-90	32	41%	87.5
1990-91	20	25%	89.2
Population	79	100%	87.3

F = 1.71, not significant at .05 level.

Appendix H

MEAN EVALUATION RATINGS, EXPERIENCE AND FORMAL EDUCATION: POPULATION

# Cases	Total Evaluation Ratings	Years of Experience	Formal Education
N = 79			
Mean	87.3	12.2	1.49
Standard Deviation	6.5	7.4	.50
Range	73-100	0-30	1.0-2.0

Appendix I

P VALUES AND DISCRIMINATION VALUES BY ITEM

BASED ON "EXCEEDS" RATINGS

Item Number	Percent "Exceeds" ratings High Group	Low Group	*p* value	D-value
1	100	12	.56	.88
2	86	8	.47	.78
3	77	42	.60	.35
4	82	12	.47	.70
5	100	29	.65	.71
6	100	25	.63	.75
7	64	21	.43	.43
8	55	4	.30	.51
9	95	42	.69	.53
10	45	13	.29	.32
11	86	33	.60	.53
12	95	21	.58	.74
13	95	25	.60	.70
14	91	29	.60	.62
15	91	21	.56	.70
16	91	63	.77	.28
17	82	17	.50	.65
18	100	25	.63	.75
19	95	21	.58	.74
20	77	12	.45	.65
21	50	0	.25	.50
22	86	21	.54	.65
23	50	17	.34	.33
24	86	38	.62	.48
25	91	4	.48	.87

Appendix J

FREQUENCY OF EVALUATION RATINGS:

HIGH-PERFORMANCE GROUP

Total Evaluation Rating	Frequency	% Total High-Performance Group
100	1	4.5
99	1	4.5
98	2	9.1
97	4	18.2
96	4	18.2
95	2	9.1
94	4	18.2
93	4	18.2
Total	22	100.0

Appendix K

FREQUENCY OF EVALUATION RATINGS:

LOW-PERFORMANCE GROUP

Total Evaluation Rating	Frequency	% Total High-Performance Group
83	5	20.8
82	5	20.8
81	2	8.3
80	3	12.5
79	3	12.5
78	1	4.2
77	2	8.3
75	2	8.3
73	1	4.2
Total	24	99.9*

* Total is less than 100.0% due to rounding.

Appendix L

MEAN EVALUATION RATINGS BY YEAR EVALUATED:

HIGH- AND LOW-PERFORMANCE GROUPS

Year Evaluated	Total Evaluations by Group: High Group	Total Evaluations by Group: Low Group	% of Group: High Group	% of Group: Low Group	Mean Evaluation Rating: High Group	Mean Evaluation Rating: Low Group
1988-89	5	9	23%	38%	96.0	79.0
1989-90	10	12	46%	50%	96.2	80.8
1990-91	7	3	32%	13%	94.7	79.7
Total	22	24	100%	100%	95.7	80.0

Appendix M

MEAN EVALUATION RATINGS, EXPERIENCE AND FORMAL EDUCATION: HIGH- AND LOW-PERFORMANCE GROUPS

	High Group = 22			Low Group = 24			Population = 79		
	Total Eval. Ratings	Years Experience	Formal Education	Total Eval. Ratings	Years Experience	Formal Education	Total Eval. Ratings	Years Experience	Formal Education
Mean	95.7	12.5	1.73	80.0	13.5	1.50	87.3	12.2	1.49
S.D.	2.1	7.2	.46	2.9	6.9	.51	6.5	7.4	.50
Range	93-100	1-25	1.0-2.0	73-83	2-23	1.0-2.0	73-100	0-30	1.0-2.0

Appendix N

EVALUATION SCORES BY SECTION: POPULATION

Section	No. Items	Maximum Score	No. Teachers	Population Range of Scores	Median Score
1	10	40	79	28-40	35
2	3	12	79	7-12	11
3	2	8	79	6-8	7
4	6	24	79	17-24	21
5	4	16	79	10-16	14

Appendix O

EVALUATION SCORES BY SECTION: HIGH- AND LOW-PERFORMANCE GROUPS

Section	No. Items	Maximum Score	High-Performance Group		Low-Performance Group	
			No. Teachers	Range of Scores	No. Teachers	Range of Scores
1	10	40	25	37-40	27	28-33
2	3	12	23	12-12	35	7-10
3	2	8	27	8-8	50	6-7
4	6	24	29	22-24	35	17-20
5	4	16	28	15-16	36	10-13

Appendix P

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY TEACHER GENDER

Gender	No. Cases	Evaluation Ratings	Experience	Formal Education
Male	41			
Mean		86.5	14.5	1.56
Standard Dev.		6.9	7.7	.50
Range		73-100	0-30	1-2
Female	38			
Mean		88.2	9.7	1.42
Standard Dev.		6.0	6.3	.50
Range		75-99	1-25	1-2
Population	79			
Mean		87.3	12.2	1.49
Standard Dev.		6.5	7.4	.50
Range		73-100	0-30	1-2

Appendix Q

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY SUBJECT TAUGHT

Subject	No. Cases	Evaluation Ratings	Experience	Formal Education
Language Arts	21			
Mean		85.7	12.3	1.38
Standard Dev.		6.8	7.0	.50
Range		75-98	2-25	1-2
Math	20			
Mean		86.9	10.8	1.45
Standard Dev.		5.2	6.7	.51
Range		80-96	2-19	1-2
Social Studies	20			
Mean		88.4	14.4	1.60
Standard Dev.		6.9	8.2	.50
Range		73-100	1-30	1-2
Science	18			
Mean		88.5	11.1	1.56
Standard Dev.		7.2	7.8	.51
Range		73-100	0-26	1-2
Population	79			
Mean		87.3	12.2	1.49
Standard Dev.		6.5	7.4	.50
Range		73-100	0-30	1-2

Appendix R

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY BUILDING

Building	No. Cases	Evaluation Ratings	Experience	Formal Education
High School	49			
Mean		86.5	13.1	1.53
Standard Dev.		6.9	7.5	.50
Range		73-100	2-26	1-2
Junior High 1	11			
Mean		86.9	8.6	1.46
Standard Dev.		5.7	7.3	.52
Range		81-97	0-21	1-2
Junior High 2	19			
Mean		89.6	11.9	1.42
Standard Dev.		5.7	7.1	.50
Range		75-97	2-30	1-2
Population	79	87.3	12.2	1.49
Standard Dev.		6.5	7.4	.50
Range		73-100	0-30	1-2

Appendix S

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY EVALUATOR

Evaluator	No. Cases	Evaluation Ratings	Teacher Experience	Teacher Formal Education
1	9			
Mean		90.2	15.6	1.78
Standard Dev.		8.0	8.0	.44
Range		78-100	2-25	1-2
2	24			
Mean		84.2	11.7	1.46
Standard Dev.		6.3	7.6	.51
Range		73-98	2-25	1-2
3	5			
Mean		87.2	17.0	1.20
Standard Dev.		4.9	2.7	.45
Range		83-93	14-19	1-2
4	11			
Mean		88.3	12.2	1.64
Standard Dev.		6.7	7.9	.51
Range		79-99	2-26	1-2
5	5			
Mean		83.6	14.8	1.60
Standard Dev.		3.6	5.4	.55
Range		81-90	7-21	1-2
6				
Mean		89.7	3.5	1.33
Standard Dev.		5.9	3.5	.52
Range		84-97	0-9	1-2
7	19			
Mean		89.6	11.9	1.42
Standard Dev.		5.7	7.1	.51
Range		75-97	2-30	1-2
Population	79			
Mean		87.3	12.2	1.49
Standard Dev.		6.5	7.4	.40
Range		73-100	0-30	1-2

Appendix T

RANKING OF EVALUATION RATINGS, EXPERIENCE,

AND FORMAL EDUCATION BY EVALUATOR

Evaluator	No. Cases	Rank in Teachers' Ratings	Rank in Teachers' Experience	Rank in Teachers' Formal Education
1	9	1	2	1
2	24	7	7	5
3	5	6	1	8
4	11	4	4.5	2
5	5	8	3	3
6	6	2	8	7
7	19	3	6	6
Population	79	5	4.5	4

Appendix U

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY GENDER: HIGH- AND LOW-PERFORMANCE GROUP

Gender	High Group = 22 No. Cases	Eval. Ratings	Experience	Formal Education	Low Group = 24 Cases	Eval. Ratings	Experience	Formal Education
Male	11				15			
Mean		95.7	12.5	1.55		79.5	15.9	1.67
S.D.		2.1	8.2	.52		2.9	6.0	.49
Range		93-100	1-25	1-2		73-83	2-23	1-2
Female	11				9			
Mean		95.6	12.6	1.91		80.8	9.4	1.22
S.D.		2.1	6.6	.30		2.8	6.8	.44
Range		93-99	2-25	1-2		75-83	2-21	1-2
Total/Grp	22				24			
Mean		95.7	12.5	1.73		80.0	13.5	1.50
S.D.		2.1	7.2	.46		2.9	6.9	.51
Range		93-100	1-25	1-2		73-83	2-23	1-2

Appendix V

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY SUBJECT TAUGHT: HIGH- AND LOW-PERFORMANCE GROUPS

	High Group = 22				Low Group = 24		
Gender	No. Cases	Eval. Ratings	Experience	Formal Education	Cases	Eval. Ratings	Experience	Formal Education
Lang. Arts	4				8			
Mean		96.0	16.3	2.00		78.9	13.6	1.38
S.D.		2.2	6.7	.00		2.9	6.7	.52
Range		93-98	9-25	2-2		75-83	5-23	1-2
Math	5				7			
Mean		94.8	12.0	1.60		82.1	9.0	1.29
S.D.		.8	7.8	.55		1.1	7.5	.49
Range		94-96	2-18	1-2		80-83	2-19	1-2
Soc.Studies	7				4			
Mean		95.4	15.6	1.71		79.0	14.8	1.50
S.D.		2.9	7.2	.49		4.2	7.3	.58
Range		93-100	5-25	1-2		73-82	2-22	1-2
Science	6				5			
Mean		96.5	7.0	1.67		79.4	18.4	2.00
S.D.		1.6	4.7	.52		2.3	2.6	.00
Range		94-99	1-13	1-2		77-83	14-21	2-2
Total/Grp	22				24			
Mean		95.7	12.5	1.73		80.0	13.5	1.50
S.D.		2.1	7.2	.46		2.9	6.9	.51
Range		93-100	1-25	1-2		73-83	2-23	1-2

Appendix W

MEAN EVALUATION RATINGS, EXPERIENCE, AND FORMAL EDUCATION BY BUILDING: HIGH- AND LOW-PERFORMANCE GROUPS

	High Group = 22				Low Group = 24		
Bldg.	No. Cases	Eval. Ratings	Experience	Formal Education	Cases	Eval. Ratings	Experience	Formal Education
1	13				18			
Mean		95.8	12.8	1.77		79.7	13.8	1.56
S.D.		2.4	8.2	.44		2.9	7.0	.51
Range		93-100	2-25	1-2		73-83	2-23	1-2
2	2				4			
Mean		97.0	5.0	2.00		82.0	15.3	1.50
S.D.		.0	5.7	.0		.8	6.2	.58
Range		97-97	1-9	2-2		81-83	2-23	1-2
3	7				2			
Mean		95.1	14.1	1.57		78.5	7.0	1.00
S.D.		1.6	4.4	.54		5.0	7.1	.00
Range		93-97	7-18	1-2		75-82	2-12	1-1
Total/Grp.	22				24			
Mean		95.7	12.5	1.73		80.0	13.5	1.50
S.D.		2.1	7.2	.46		2.9	6.9	.51
Range		93-100	1-25	1-2		73-83	2-23	1-2

Appendix X

NUMBER OF CASES AND MEAN RATINGS BY EVALUATOR: HIGH- AND LOW-PERFORMANCE GROUPS

Evaluator	Building	No. Cases			Mean Ratings		
		High Group	Low Group	Population	High Group	Low Group	Population
1	HS	5	2	9	96.6	80.0	90.2
2	HS	3	11	24	95.7	78.7	84.2
3	HS	1	2	5	93.0	83.0	87.2
4	HS	4	3	11	95.5	80.7	88.3
5	JH1	0	4	5	--	82.0	83.6
6	JH1	2	0	6	97.0	--	89.7
7	JH2	7	2	19	95.1	78.5	89.6
Total		22	24	79	95.7	80.0	87.3